IDA PAPER P-2317

FURTHER EXPLORATIONS IN ESTIMATING THE MILITARY VALUE OF TRAINING

Seymour J. Deitchman

January 1990

Prepared for
Office of the Under Secretary of Defense for Acquisition

IDA
INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street, Alexandria, Virginia 22311-1772

DEFINITIONS

IDA publishes the following documents to report the results of its work.

Reports

Reports are the most authoritative and most carefully considered products IDA publishes. They normally embody results of major projects which (a) have a direct bearing on decisions affecting major programs, (b) address issues of significant concern to the Executive Branch, the Congress and/or the public, or (c) address issues that have significant economic implications. IDA Reports are reviewed by outside panels of experts to ensure their high quality and relevance to the problems studied, and they are released by the President of IDA.

Group Reports

Group Reports record the findings and results of IDA established working groups and panels composed of senior individuals addressing major issues which otherwise would be the subject of an IDA Report. IDA Group Reports are reviewed by the senior individuals responsible for the project and others as selected by IDA to ensure their high quality and relevance to the problems studied, and are released by the President of IDA.

Papers

Papers, also authoritative and carefully considered products of IDA, address studies that are narrower in scope than those covered in Reports. IDA Papers are reviewed to ensure that they meet the high standards expected of refereed papers in professional journals or formal Agency reports.

Documents

IDA Documents are used for the convenience of the sponsors or the analysts (a) to record substantive work done in quick reaction studies, (b) to record the proceedings of conferences and meetings, (c) to make available preliminary and tentative results of analyses, (d) to record data developed in the course of an investigation, or (e) to forward information that is essentially unanalyzed and unevaluated. The review of IDA Documents is suited to their content and intended use.

The work reported in this document was conducted under contract MDA 903 89 C 0003 for the Department of Defense. The publication of this IDA document does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that Agency.

This Paper has been reviewed by IDA to assure that it meets the high standards of thoroughness, objectivity, and appropriate analytical methodology and that the results, conclusions and recommendations are properly supported by the material presented.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
January 1990

3. REPORT TYPE AND DATES COVERED
Final--March 1989 to December 1989

4. TITLE AND SUBTITLE
Further Explorations in Estimating the Military Value of Training

5. FUNDING NUMBERS
C - MDA 903 84 C 0031
T - T-D2-435

6. AUTHOR(S)
Seymour J. Deitchman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Paper P-2317

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
OUSD(A)/R&AT
The Pentagon, Room 3D129
Washington, DC 20301-3080

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report describes the continuation of work started earlier to find ways of quantifying the military value of training. The earlier work used a large-scale simulation model of warfare to examine the potential effects of assumed force improvements imputed to training on the outcome of a postulated war in Europe. The work described in this report gathered some of the available data showing the effects of training on force effectiveness, and it estimated the cost of training, in the areas of platoon-size armored combat and bombing accuracy by tactical attack aircraft. These results were compared with the results of analyses describing the effects of equipment improvement in the same areas of unit combat. The report shows the size of the effects in each case, and it evaluates the relative contributions of training and hardware advances to improvement of force effectiveness and the relative costs of the two approaches. Conclusions are reached about preferred approaches to such evaluations, and desirable future elaborations of the research are outlined.

14. SUBJECT TERMS
Military Value, Training, Force Effectiveness, Training Cost, Armored Unit Combat, Bombing Accuracy

15. NUMBER OF PAGES
68

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

IDA PAPER P-2317

FURTHER EXPLORATIONS IN ESTIMATING THE MILITARY VALUE OF TRAINING

Seymour J. Deitchman

January 1990

INSTITUTE FOR DEFENSE ANALYSES

Contract MDA 903 89 C 0003
Task T-D2-435

PREFACE

The work described in this report was performed under Task Order T-D2-435, Contract MDA903-84-C-0031. The report was reviewed in draft by Dr. Earl A. Alluisi, Dr. Dexter Fletcher, Mr. Stanley Horowitz, Dr. Jesse Orlansky, and Mr. Joseph Ruzzi, and their many suggestions for improving the report were most helpful. Final responsibility for the report's analysis, discussion and conclusions rests with the author.

ACKNOWLEDGMENTS

Mr. Joseph Ruzzi of the IDA Science and Technology Division staff performed all the assembly, statistical analysis and preparation for the presentation of the data from the SIMNET IVIS tests that are described in this report. His help in putting the data into form for interpretation, and in setting up computer programs that facilitated manipulation and exploration of the significance of the data, was invaluable.

Mr. Paul Monday, BBN Corporation, who assembled the output data from the IVIS tests, performed intervisibility calculations, and transmitted the data and the results of the calculations to IDA in hard copy and in machine readable form also made an essential contribution to this work.

Finally, the inputs of Mr. Stanley Horowitz of the IDA Cost Analysis and Research Division, who made available the data on the effects of flight hours on tactical bombing accuracy, and Mr. Joseph Stahl of that Division, who assembled much of the cost data used, are gratefully acknowledged.

ABSTRACT

This report describes the continuation of work started earlier to find ways of quantifying the military value of training. The earlier work used a large-scale simulation model of warfare to examine the potential effects of assumed force improvements imputed to training on the outcome of a postulated war in Europe. The work described in this report gathered some of the available data showing the effects of training on force effectiveness, and it estimated the cost of the training, in the areas of platoon-size armored combat and bombing accuracy by tactical attack aircraft. These results were compared with the results of analyses describing the effects of equipment improvement in the same areas of unit combat. The report shows the size of the effects in each case, and it evaluates the relative contributions of training and hardware advances to improvement of force effectiveness, and the relative costs of the two approaches. Conclusions are reached about preferred approaches to such evaluations, and desirable future elaborations of the research are outlined.

CONTENTS

FIGURES

TABLES

ABBREVIATIONS

IVIS	Intervehicular Information System
JMEM	Joint Munitions Effectiveness Manual
MILES	Multiple Integrated Laser Evaluation System
MOE	Measures of Effectiveness
NTC	National Training Center
SIMNET	Simulation Network
WCI	Weighted Casualty Index

EXECUTIVE SUMMARY

This report is the second in a series designed to explore and, if possible, quantify the military value of unit training. The first report (Deitchman, 1988) used a warfare simulation model to examine what would happen to the outcome of a war if unit training could lead to various assumed increases in the capability of armored ground forces and tactical air attack forces to destroy the enemy and to survive themselves. It was found that factors-of-two improvements in parameters considered relevant to training could reverse the course of a war in Central Europe, as portrayed by the model. Central, unanswered questions inherent in this approach were whether the assumed factors of improvement due to training could be achieved, how their achievement could be quantified, and whether the costs to achieve them could be estimated. If they could, then it would be possible to trade off training costs against the costs of new hardware to achieve similar results, and spend resources at the margin in the most effective way.

This report explores some of these critical questions. Specifically, it examines a sample of available data from various sources, on tank warfare at platoon level and on tactical bombing at squadron level, to see whether and how the data might answer the following two questions:

1. Can realistic, quantitative values for unit training effectiveness be determined, to lend credibility to model-based calculation of the military value of training expenditures; and
2. Is it feasible to trade off expenditures for training against those for hardware to improve force capability?

A subsidiary question was to compare results from training exercises with the SIMNET network of armored vehicle simulators with results from field exercises, both simulating units in armored combat. This would shed some light on what could be learned from both approaches in controlled unit training experiments, and on the relative costs of the two.

The results of the explorations are summarized in the table on the next page. Their significance for the larger questions considered is explored in detail in Chapter IV.

Table S-1. Overall Summary of Training/Equipment Comparisons

WARFARE AREA	COMPARISON MADE	RESULTS: EFFECTIVENESS	RESULTS: COST	REMARKS
Armored combat by platoon-sized forces	Field training using M-60 tanks with SIMNET exercise using M-1 tanks, in three-week exercises for each (1)	• Improvement in tank MOE in field tests equivocal, depends on specific conditions and units operating • Offense improved more than defense in simulator tests, with single unit operating	Three-week field training exercise costs about 30 times as much as 3-week SIMNETexercise; but note: test facility costs not included because imbedded in Army training facilities budget	• SIMNET exercise statistics too meager to give credible trend in exchange ratio • Field test platoon had combined arms augmentation; SIMNET platoon, all tanks; different organization affected relative outcomes
	Field training using M-60 tanks with M-60 upgrade from A1 to A3 model (2)	• 35% overall improvement in Blue's favor over 3-week training period • 20% improvement in Blue's favor with hardware upgrade	• $2.8 million per platoon for tank upgrade • $2.25 million for training if once/yr; $9 million if 4 times/yr	• Field training with combined arms platoon; hardware analysis with all tank platoon • All dollars are 10-year costs
Tactical bombing by fighter-bomber aircraft	A-10 and F-16 squadron improvement with monthly flying hours (3)	• Accuracy increases by factor of 1.8 when flight hours increase from 10 to 40/month, for pilots with>1400 career hrs.	~$650 million for squadron flight hrs over 15 yrs, including Sqdn fixed base costs; ~$130 M if not incl	• Similar results and costs for both aircraft • Accuracy overall depends 60% on career flt hrs, 40% on current flt hrs
	A-10 with F-16, squadron re-equipment (4)	• Accuracy increases by factor of 2, regard-less of flight hours	~$800 million for squadron A/C upgrade, 15-yr cost	• Comparison is for day-visual dive bombing; F-16 is multirole A/C, A-10 is not
	A-7 with F/A-18 bombing accuracy (5)	• F/A 18 accuracy 15% - 75% better than A-7, across wide range of bombing parameters • A-7 pilots would require 1125 more career hours to achieve 75% average accuracy improvement (High-end estimate)	• $17 million/sqdn if F/A-18 bombing system can be retrofit to A-7 • $27 million/sqdn for upgrade by 1125 career hrs per pilot • $480 million/sqdn to change from A-7 to F/A-18	• A-7 avionics cost and training upgrade are one-time costs; change of A/C is 15-yr cost • Not known whether A-7 upgrade to F/A-18 bombing system would be feasible or lead to same results

SOURCES:
1. ARI, 1978; BBN, 1989 for effectiveness; IDA- developed data (from Army Forces Command) for costs
2. ARI, 1978; Graves, 1974 for effectiveness; Graves, 1974 and IDA- developed data for costs
3. Cedel and Fuchs, 1986 for effectiveness; IDA- developed data for costs
4. Cedel and Fuchs, 1986 for effectiveness; IDA- developed data for costs
5. Mairs, 1986; JMEM for effectiveness; IDA- developed data for costs
6. Unclassified cost data: for hardware,IDA-developed from OSD sources; training cost data furnished by U.S. Army Forces Command

The following conclusions have been drawn from this work about quantifying the military value of training for system acquisition and resource allocation decision purposes:

1. It appears to be possible to quantify the military value of training, but data for the purpose will best be gathered through explicitly designed trials. However, resource availability will dictate that ad hoc opportunities to take advantage of existing data should not be foregone.
2. The cases explored suggest that either training or equipment improvement for specific military tasks improve force effectiveness by roughly comparable magnitudes. Depending on the cost elements included, training or equipment improvement may be either comparable in cost or else training tends to cost less--sometimes considerably less. (Other cases than those examined here may show a different balance.)
3. Elements of training and equipment improvement and replacement will have to be combined to have any chance of achieving the improvement in unit capability that the explorations with the TACWAR model indicated would be necessary to reverse the course of the NATO Central Region war modeled. Factors of two improvements in such skills as killing tanks, bombing targets and survival were required in the model. The actual achievements that were found in the available data and analyses, due either to training or hardware improvement, ranged from 20 percent to 35 percent for armored combat, and from 15 percent to 100 percent in bombing accuracy.
4. Some of the data suggest that more automatic modes in new systems may reduce the requirement for individual proficiency training, freeing resources for more unit training.
5. Experimental data about the impact of training on unit effectiveness gathered under controlled conditions in simulator networks like SIMNET will be useful in quantifying the military value of training, and they will be better controlled and less expensive than field exercises.
6. Regardless of how the cost and performance comparisons may vary when explored in more detail, it is apparent from the magnitudes of the costs and effects that algorithms for allocation of resources among training, equipment improvement and force size must be devised to seek the most efficient resource allocation among the available approaches to force improvement. Such algorithms are not currently used in cost-effectiveness analyses of new systems.

These results must be taken as suggestive rather than conclusive, because none of the data were generated for purposes such as those of this report. Therefore, in none of the comparisons is any one case exactly comparable to any other, nor are the bases of comparison--the conditions under which the performance and cost data were generated, the experimental or analytic designs, the environments, the scenarios, or the purposes for which the data were generated--the same for any two sets of data or analytical results. And yet, each comparison overlaps with at least one of the others, so that there is some check on the size and direction of each of the answers via redundancy, within the boundaries permitted by the data.

Thus, the results do lead to new insights about the analytic process being explored, they shed some light on the answers to the research questions examined, and they are encouraging in confirming future directions for inquiry.

It is recommended that future work continue in the same combat areas (small-unit armored combat and tactical air attack of ground targets), because of their importance and the potential availability of data. The next steps should include:

- A much more thorough data exploration (a) to define better the nature of the experimental data available, and (b) to see what more can be learned about the effects of exercises and training on unit performance;
- Enlisting the Services' interest and help in designing and carrying out relevant trials at SIMNET and available, analogous USAF and USN simulator complexes, to shed light on some of the important issues the present work has raised.
- Exploring whether and how exercise and simulation data gathered at low levels of military organization, such as platoon or flight levels, aggregate to describe performance of units at higher levels of organization, such as battalion or squadron, or higher, levels.
- Using the data gathered above to devise resource allocation algorithms incorporating both training and equipment effectiveness parameters.
- Experimenting with the algorithms using warfare simulation models such as are used by the DoD for budget planning and evaluation purposes, to explore the algorithms' ranges of utility and how they might affect resource allocation in the Department of Defense. The results should be subjected to military judgment, to ascertain whether the aggregation from organization levels such as platoons and battalions to divisions and armies appears to give reasonable results.

I. INTRODUCTION

Both training and hardware can be subjected to the same evaluation criteria in budget planning: increased capability for the money spent, or, in the usual measurement terms for defense planning, criteria of cost-effectiveness. Subjecting them to evaluation by such criteria is essentially the same as assigning a "military value" to each, training or hardware. The most direct way to assess the military value of training or hardware after a given expenditure is through observation and measurement of performance in combat. This carries sometimes high risks, and is not always possible. Therefore, many surrogates have been developed for the purpose.

The evaluation techniques for training and for hardware differ greatly. There are well-known analytical, simulation, and field testing techniques in the hardware area to ascertain some measures of military value. These techniques are not always completely valid or applicable, but they have been developed over a period of decades and defense planners tend to have confidence in them, bred of familiarity. In the training area, the military value of training can best be assessed using unit rather than individual performance, even though unit performance starts with individual performance; forces operate in units most of the time. Unit performance assessment depends on measurement before, during, and after training. Whether the results of such assessment derive from combat or from field exercises and related activities, there is extensive qualitative military judgment about the effects of training on unit proficiency, most of it positive. But there is little quantitative data to shed light on the effectiveness of training in a way that would permit assessments of military value of training with the same degree of confidence accorded to such assessments in the hardware area.

This is the second report about a series of explorations intended to find ways to quantify the military value of unit training in a form useful for decisions about expenditure of resources. The first report dealt with the question of what the military value of unit training might be if the training resulted in certain assumed levels of improved force proficiency. To examine that question a large-scale simulation model of land/air warfare was used. Values of key weapon system performance parameters were changed arbitrarily to simulate changes in capability that might be attributed to unit training, and the question was asked whether those changes would make any difference in the outcome of a war. The answer was that they would, but that the degree to which the assumed parameter changes reflected the actual effects of training remained to be determined.

Specifically, the IDA TACWAR model was used to describe the outcome of a postulated war between NATO and Warsaw Pact (WP) forces in the NATO Central Region. The TACWAR model is a computer simulation of theater-level warfare, adapted to NATO Central Region defense (but not exclusively). It has separate corps sectors, includes ground and air warfare with combined arms, and portrays rear area as well as front line operations (Kerlin, 1977).

In the base case, which represented the NATO and WP forces and their capabilities as they might be in the early 1990s without any arms control agreements, the NATO forces lost the war by a significant margin, as measured by the average westward movement of the forward edge of the battle area (FEBA). Many parameter changes assumed to represent the effects of unit training were explored. Of specific interest in this report, if it were assumed that NATO's armored forces, through training, could double their ability to destroy enemy armor and to keep from being destroyed themselves, or that NATO's air-to-ground tactical air forces could double their ability to destroy enemy targets on the ground and to keep from being shot down by opposing air defenses (i.e., a factor of four improvement in overall capability, in each case), then either of those improvements in (armor or tactical air) performance would be enough to reverse the course of the war as described by the TACWAR model. In the armor case it was enough to increase target killing capability; attack aviation required both improved target killing capability and ability to evade air defenses.

It was shown that the details of the differences between the armor and tactical air outcomes were attributable to differences in how the model treated the systems in its system and force interaction algorithms. More fundamentally, the question was raised of whether the parameter changes assumed to represent the effects of improved training could actually occur. If actual parameter values describing training effects on force capability can be determined, these values could be used in evaluation models like TACWAR, or others of the many available, to assess the impact of training expenditures on the outcomes of battles or wars, and to compare the value of expenditures at the margin for training or for new hardware acquisition for improving force capability.

The further explorations described in this report consisted of:

1. Examining the literature for prior results of training exercises in the two areas noted above--armored warfare and ground attack by tactical aircraft, which were chosen because they seemed to offer the most readily available sources of data on the effects of training and equipment improvement;
2. Examining the results of the training exercises in the DARPA-sponsored SIMNET armored warfare training network of manned tank simulators at Fort Knox, Kentucky;
3. Comparing the effectiveness improvements resulting from the training exercises with those from improved hardware, using sources reflecting reasonable estimates of the gains that could be expected from the hardware;
4. Estimating costs for the training levels that led to the effectiveness changes experienced and the costs of the hardware changes examined;
5. Comparing the costs and effectiveness of the various approaches to improving unit effectiveness.

II. RESEARCH QUESTIONS, APPROACH AND DATA SOURCES

A. RESEARCH QUESTIONS

Two primary questions were examined in this part of what is intended to be a continuing series of studies:

1. Can realistic, quantitative values for unit training effectiveness be determined that would lend credibility to model-based calculation of the military value of training expenditures; and

2. Is it feasible to trade off expenditures for training against those for hardware designed to achieve similar effects in combat, i.e., to improve force capability?

The exploration of these questions in this study was designed mainly to test the feasibility of finding or generating data to shed light on them, and to test whether these apparently straightforward research questions were in fact amenable to quantitative analytic treatment. For such a purpose, already existing data were sought, since without the "proof of principle," as it were, there would be slim justification for more elaborate and expensive experimental gathering of new data. Given success in this preliminary phase, a more detailed and rigorous data-gathering phase could be planned.

Observation of the change in unit performance before and after appropriate training and subsequent combat is associated with wartime operations that do not lend themselves to rigorous measurement or gathering of experimental-quality data. Although the opinions of experienced commanders could be sought (as was recommended in Deitchman, 1988), such data would be qualitative rather than quantitative. Qualitative inputs will, in the long run, be necessary to confirm the "reasonableness" of quantitative data gathered under other than real battle conditions, but it would be analytically more appealing to have a solid quantitative base from which such qualitative judgments could start.

Other means for gathering the appropriate data would include field measurement of unit performance in two-sided exercises and manned laboratory simulations of combat. Both techniques have been developed over recent years to the point where each provides useful quantitative data on unit performance.

Exercise data are now being gathered at training ranges such as the National Training Center armored warfare training ground. This was preceded historically by training exercises in the field, such as those called called REALTRAIN in Europe, some of the results of which were available for analysis here. The NTC gathers data in mock combat using actual equipment through the Multiple Integrated Laser Evaluation System (MILES) and a position reporting system for individual vehicles, which together measure who shot at and hit which target, so that casualties can be measured together with observation of the positional outcome of the mock combat. The Air Force conducts such activities as close air support in operations called Red Flag, in the Nevada desert. The SIMNET developed by DARPA and the Army, a network of manned training simulators at Fort Knox, Kentucky and elsewhere (Orlansky and Thorpe, in publication), offers a way to observe troop performance in simulated training conditions including armor, artillery and close air support, and to have a start-to-finish record of all the events in a unit engagement, at significantly lower cost than field exercises and in circumstances where the engagement parameters can be better controlled than in the field.

Each of these techniques varies different parameters of the training process and unit performance, and allows gathering of different kinds of data that explore diverse aspects of the exercises. But the measurements do overlap. Thus, a subordinate research question is:

3. To explore the correspondence between the results of field training and the manned simulator network, to examine how each illuminates the first two questions.

B. OVERVIEW OF COMPARISONS

Table 1 on the following page summarizes the comparisons made. Specific passages have been underlined for emphasis to remind the reader of the research questions for which answers were being sought.

The subsequent sections of this chapter will discuss further details of the data sources and the way they were used in making the comparisons shown.

Table 1. Comparisons Among Training, Simulation and Hardware That Were Made During This Analysis

PURPOSE OF COMPARISON	WHAT WAS COMPARED
Compare field test with simulation in describing armored warfare training effects	REALTRAIN field training with SIMNET IVIS exercise, both at platoon level
• Assess improvements achievable in tank platoon combat capability through training or equipment improvement • Assess relative cost-effectiveness of training and relevant hardware improvement in tanks, in terms of unit combat performance at platoon level	• REALTRAIN, IVIS, and M-60A1/M-60A3 comparative evaluation • REALTRAIN field training with M-60A1/M-60A3 comparative evaluation
• Assess improvements achievable in bombing accuracy due to training or relevant equipment improvement • Assess relative cost-effectiveness of training and relevant equipment improvements on attack aircraft bombing accuracy	• A-7 training results and A-7//F/A-18 bombing system accuracies and • A-10 and F-16 training results (Both apply to both questions)
Ascertain effect of automatic machinery on extent of training needed (serendipitous result)	M-1 tank crew selection and performance tests, and F/A-18 and AV-8B bombing data

C. IMPROVEMENTS IN ARMORED UNIT CAPABILITY DUE TO TRAINING

Two sets of data were used to shed light on the questions associated with unit training in armored warfare. The first resulted from a 1975 field exercise in Europe designed for U.S. Army platoon level armored unit training, called REALTRAIN (ARI, 1976). This was described in the reference as the first realistic, unit-on-unit training exercise in a variety of terrains and tactical situations, as distinct from gunnery training and similar range training to sharpen individual skills. The combat units consisted of tank platoons, each reinforced by a TOW (heavy, vehicle-mounted antitank weapon) section and two infantry squads in armored personnel carriers (APCs). The reference does not give the exact numbers of individual systems; they are presumed to be 4 to 5 tanks, two TOW launchers; and two M113 APCs with 24 infantry soldiers (the numbers of infantry were given). From the date of the exercise, the tanks are assumed to be M-60A1s. Offense, defense and meeting engagements in varying terrain were part of the exercise for platoons opposing each other. The progression of platoon capability with time as the training exercise continued for its three week duration was of interest in the current context.

The combat simulation results from SIMNET were also at platoon level, taken over a three-week period (Headquarters, Department of the Army, 1988). They consisted of a series of tests, conducted in the summer of 1988, involving five manned tank simulators (a 4-tank platoon and a company commander), in a series of engagements with a comparably-sized threat involving tanks and BMPs (Soviet armored infantry fighting vehicles) operating in a semiautomated mode (i.e., they could be maneuvered by an "authority" outside the simulation, and they could fire when appropriate relationships to opposing targets obtained, but they were not represented by manned simulators on the "battlefield"). Artillery effects were also simulated. The Blue tanks simulated in this case were M-1 tanks with "Block II" equipment. The Block II equipment included an improved navigation system, improved night vision system, and an Intervehicular Information System (IVIS) from which the test series took its name.

The IVIS data were, as will be shown later, similar in character to the REALTRAIN data. The analysis was limited to observation of trends in the Blue platoon performance over the three-week test series with the Block II equipment; the base case (no Block II equipment) and three equipment familiarization sessions were deleted. Thus, even though the tests were conducted for a different purpose than training (evaluating the effects of the "new" equipment on unit performance) any trends in unit performance with time using the same equipment could be taken as a first approximation to the effects of learning, and they were therefore considered to be analogous to the results of an exercise conducted as though it were for training. Offensive and defensive exercises in a specific terrain area were run; there was no exact parallel to the meeting engagements of REALTRAIN.

The REALTRAIN and IVIS results were compared with each other to see how closely the results might correspond, and to assess the cost differences for two exercises of roughly the same size, one in the field and one using the simulator network. The REALTRAIN exercise results, which were based on unit performance using the M-60 tank, were compared with modeled effects of equipment changes in the M-60 tank (described below) to assess the relative effects of training and equipment in a restricted set of armored warfare situations.

D. IMPROVEMENTS IN ATTACK AIRCRAFT ACCURACY DUE TO TRAINING

Although bombing practice is a regular part of attack pilot training, data describing the effects of unit level training on bombing accuracy are not extremely plentiful. Two references provided useful data in this area: a 1986 analysis of training and experience on the bombing accuracy performance of Navy pilots flying A-7E aircraft (Mairs, et al., 1986), and a 1986 USAF report analyzing similar performance data for A-10 and F-16 squadrons (Cedel and Fuchs, 1986).

The A-7 report provides a statistical analysis of the bombing accuracy of individual pilots in A-7 squadrons (but not squadron performance) as a function of career flight hours and career jet hours. The USAF report describes patterns of squadron average miss distance,[1] as a function of both career flight hours and of recent flight training activity. It also presents a model derived from the data permitting separation of the effects of long term and recent experience. A further analysis by Hammon and Horowitz of IDA clarifies the distinction (Hammon and Horowitz, 1989). Hammon and Horowitz also show some data that bear on day-visual bombing performance with the F/A-18 and AV-8B aircraft in the manual and the automatic modes, relevant to the effects of flying hours on bombing performance. It should be noted that no data were included in the available references indicating how much bombing practice was included in the career hours. A related analysis

---

[1] Squadron average miss distance is related but not identical to Circular Error Probable, CEP, the radius of the circle within which half of the bombs would drop; the reference does not indicate the exact relationship that was used.

of skill acquisition curves (Lane, 1986) suggests that it may be reasonable to assume that bombing accuracy will improve together with all other skill improvements, at least to some possibly asymptotic level, as career flying hours increase.

The A-7 data were used to compare the effects on bombing accuracy of training in the A-7 and of improving the bombing system from the level of the A-7 to that of the F/A-18 aircraft. The USAF report on the A-10 and F-16 contains enough comparative data to permit separation of the effects of equipment improvement from those of training with some confidence, and this report was also used in the training/equipment comparisons.

E. EFFECTS OF AUTOMATIC EQUIPMENT ON TRAINING REQUIREMENTS

An Army report on variation of training effectiveness with soldiers' grades on mental capability tests (U.S. Military Academy, 1984) sheds some light on the effect of automatic equipment in tanks on training requirements for specific levels of proficiency. Together with the aircraft data comparing manual with automatic bombing, noted above, these data bear on the variation of training requirements with sophistication levels of equipment technology--the data are skimpy, but of sufficient interest and relevance to contribute to the results in this report.

F. EFFECT OF EQUIPMENT IMPROVEMENT ON UNIT PERFORMANCE

Analytical data describing effectiveness changes with equipment changes in tanks and aircraft were sought, to enable comparisons of equipment-improvement and training effects.

Analyses of equipment improvements tend to be classified when they deal with the detailed specifications of the equipment and its performance under specified conditions. In order to keep this report unclassified, the effects of equipment changes are described in terms of percentage changes and broad performance boundaries, rather than specific values of performance parameters or the case-by-case results of model analyses. Since, as noted above, the equipment analyses tend to describe what are actually stochastic phenomena in deterministic terms, and since the training outputs are described in stochastic terms, it is believed that not much in the way of useful comparison is lost by the necessary "masking" of the details of the equipment analyses.

For the tanks, a 1974 IDA analysis of main battle tanks showed the calculated effect of improving the M-60 tank from the A1 to the A3 configuration (Graves, et al., 1974). The most important hardware changes, from the point of view of crew proficiency, were an improved fire control computer and a stabilized main gun (e.g., the effects of improved armor or ammunition, if they were part of the tank improvement, would not be related to crew proficiency in any direct way). The output measures of the Tank Exchange Model used in the analysis employed both offense and defense scenarios in a variety of terrains in platoon-sized engagements, and thus could be compared with the REALTRAIN output measures.

For the attack aircraft, two sources provided descriptions of bombing accuracy with different bombing system generations. One provided a comparison of the A-7 with the F/A-18 aircraft--both single-seat attack bombers that could attack in the same modes with different capability (JMEM). The other was the same USAF report cited in the training case, above (Cedel and Fuchs, 1986); it also showed extensive data describing the difference in bombing accuracy with the A-10 and F-16 aircraft for the entire range of pilot lifetime flight hours that the report covered in its detailed analyses.

G. COSTS

Although costs were obtained from a variety of sources they were kept comparable in specific parts of the analysis.[2]

For field training in armored warfare, data describing the costs of an average brigade-sized exercise at the NTC were scaled down proportionately to describe the costs of exercises involving a unit the size of the reinforced platoon in REALTRAIN. The major available costs of the field exercises at NTC included those involved in bringing the units' personnel and equipment to the test range and the cost of operating the units there. Troop pay and allowances were not included, based on the rationale that the troops are paid whether they are involved in formal training or not. The average cost for a three-week, brigade-size exercise at NTC, involving about 4400 men, about 900 vehicles and 46 helicopters is about $5 million. The range support costs are imbedded in larger Army

2 Data on the costs of the A-7 and F/A-18 bombing systems, on aircraft flight hour costs, and on the costs of NTC exercises were assembled by J. Stahl of IDA. The training costs were furnished by the J3, Resource Management Office, Department of the Army, Headquarters Forces Command. Life cycle costs for the various aircraft had been assembled for the author by M. Olver of IDA for a prior analysis and were used in the present case.

training facilities line items and could not be separated out for this analysis. They were therefore not included in the per-person cost of the exercise.

The costs of the platoon-level REALTRAIN exercise were taken as 4-1/2 percent of the cost of an average brigade-size exercise at NTC, because about 4-1/2 percent of the numbers of troops and vehicles were estimated to be involved. This estimate is crude, but there was no information available as to the nature of the REALTRAIN test range in Europe or its cost structure. The bulk of the NTC exercise costs are in transportation, and although distances are shorter in Europe, the bulk of the transportation costs are involved in loading and unloading at the terminals. The major problem is that the allocated costs of range operation for a small unit are also not known.

The cost of running the SIMNET exercise was taken as the same as the cost of transporting personnel, only, to an NTC exercise, *per military person involved*, also excluding pay and allowances. The estimated numbers of people were taken as about the same as the estimated numbers of REALTRAIN tank crew members per team, about 30 people. As in the case of the NTC range, the cost of operating the SIMNET facility, itself, is imbedded in other Army line items and could not be identified for this analysis. Therefore, the SIMNET operating cost was not included in the cost of the IVIS exercise.

Failure to include range and facility operating costs in the cost of the training exercises is consistent between the two kinds of training exercise included in the analysis, but it must be remembered that the omission could distort any cost comparisons with hardware by some uncertain amounts. True training costs for ground force units will always be higher than those shown.

The cost differences between the M-60A1 and the M-60A3 tanks were obtained from Graves, et al. (1974).

The aircraft cost comparisons made were such that total costs for pilot career training were not needed; incremental costing was sufficient. Bombing system and total aircraft system costs were applied to the A-7/F-18 comparison, to illuminate the difference between retrofitting an improved bombing system to an existing aircraft (the A-7) and replacing the aircraft system entirely, while total aircraft system costs only were applied to the A-10/F-16 comparison.

Costs per flight hour for the aircraft in question were used for training costs of attack pilots. The costs used were total cost per flying hour including fixed and variable parts. The fixed parts are base overhead costs associated with the presence of the aircraft whether they fly or not. The variable parts are crew, fuel and some maintenance costs that vary with flying hours. The variable costs are about 20 percent of the total. However, the division is somewhat arbitrary, since even the fixed costs will vary with flying hours to the extent that increasing flying hours bring the aircraft to the end of service life (ESL) more rapidly. Thus the true cost per flying hour is between the variable cost and the total cost that was used. For convenience, since the costs were quite close, a variable cost per flying hour of $1000 was used for all aircraft. For the same reason, A-10 and A-7 total cost per flight hour were assumed to be the same since the comparisons made would not be sensitive to any difference.

In the comparisons between hardware and training costs in bombing system replacement, only the variable flight-hour costs were used for the A-7, on the assumption that the improved bombing system could be retrofit to that aircraft and the fixed flying-hour costs would not be changed enough by that upgrade alone to affect the cost comparison significantly within the level of approximation of the overall analysis. The hardware upgrade for comparison was taken to be a one-time cost. For the cases where hardware upgrades in the form of total aircraft system replacement were being compared with training (the F-16 replacing the A-10 and the F/A-18 replacing the A-7) the total flight-hour cost was used for training cost in the comparisons. This would include support cost, and would therefore be roughly comparable with the aircraft 15-year life-cycle cost ascribed to the new aircraft. It is believed that system and budget planners would take similar approaches in the respective cases.

In all cases where cost differences or absolute costs were compared, the cost data available were normalized to FY1989 dollars using the DoD annual inflation factors.[3] None of the multi-year cost streams was discounted.

H. DATA QUALITY

The following discussion of the qualities of the performance and cost data used for individual comparisons shows factors in the data that will have an uncertain effect on the results of the comparisons, and should therefore be borne in mind.

The platoon-sized armored forces in the REALTRAIN field training exercise were not the same as those in the available SIMNET exercise, and the similarly sized forces evaluated in analytical models to compare equipment improvements with training are

---

3 Available from OSD to the IDA Cost Analysis and Research Division.

different from either of the experimental forces. The tanks that were used in the two experimental cases were not the same. And, finally, the scenarios, terrain and other physical factors that can affect the outcomes were not the same in any of the data sets compared--experimental or analytical.

Capabilities achievable through training approach asymptotic values as training proceeds over long enough periods. The tactical bombing training data used here did cover long enough periods to reach the asymptotic value, but the armored combat exercises lasted only three weeks, which is probably not enough time to approach that value. On the other hand, long-term career experience probably also affects the rate at which the asymptotic value is approached in armored combat, as it does in tactical bombing. Nothing is known about the career experience of the personnel involved in the armor training exercises described in this study.

Similar parameter differences to those in the armor case existed in the attack bombing cases. It is not known whether the A-7 aircraft used in the experimental measurement of the variation of bombing accuracy with experience (Mairs, et al., 1986) were configured in the same way as the aircraft for which bombing accuracy data were available in "handbook" form (JMEM). The comparison of A-7 and F/A-18 bombing systems assumes that the F/A-18 bombing system would perform the same way in the A-7 (if it could be retrofit to that aircraft) as it does in the F/A-18, but in the actual case the differences in aircraft flying qualities and even cockpit layout would also affect bombing accuracy of the two aircraft in ways that are not identified here.

The experimental data points are widely scattered in all cases. Sample sizes for the armored warfare comparisons are small, and although they are larger for the bombing accuracy cases the pilots in some of the cases were drawn from single squadrons rather than at random from the entire pilot pool. Trends in the effects sought thus remain indicative of tendencies in stochastic processes and cannot yet be considered definitive in all the cases examined.

The analytical results evaluating hardware performance through tank battle exchange ratios and bombing system accuracy come from deterministic (rather than Monte Carlo) models that are not statistical in nature. However, those models describe what are essentially stochastic processes with an implication of precision that they cannot have. Thus, strictly speaking, the experimental training and the analytical equipment evaluation results examined must be considered only partially comparable.

The cost data include similar uncertainties. In the tank warfare cases, it was impossible to find the range or training facility cost that should be ascribed to the exercises being examined. In the air warfare cases, it is uncertain whether the total relevant costs for training and for equipment are being incorporated in the comparisons in a fully consistent manner. Some training costs tend to be hidden, while equipment costs tend to include total life-cycle cost according to well-known cost estimating and accounting procedures.

III. RESULTS OF SEPARATE EXPLORATIONS

The following discussion will describe the essential, relevant elements of the experiments and the analyses, the measures of effectiveness (MOE) of interest in the present context, and the significant results in each case. Intermediate data are described to the extent necessary to interpret the results.

A. TRAINING

1. REALTRAIN

In the REALTRAIN exercises, one team (reinforced armor platoon), called Team A, held the field for three weeks, while three other teams, each successively dubbed Team B, were brought in for a week at a time to fight against them. The teams took turns in engaging in offensive or defensive combat, and there were also meeting engagements. The exercises were divided roughly into offense and defense, for half of the engagements, and meeting engagements, for the other half of the engagements. There were 54 two-sided exercises in all.

The measures of effectiveness of interest to this analysis (there were others of similar character) included tank casualties on each side, and a weighted casualty index (WCI) that added together all casualties of all kinds of vehicles, and infantry, for each side, weighting the numbers of casualties of vehicles or troops, as the case might be, according to a judgmental weighting factor expressing the importance of each casualty element in the view of the rater (e.g., individual tanks were weighted 35, and individual infantry soldiers were weighted 1).

Table 2 (taken from ARI, 1976) shows the overall results of the exercises for Teams A and B, in terms of the WCI for each team at the end of each week. These results wrap up all kinds of combat (offense, defense, meeting engagement) in all kinds of terrain in the single output number, which was found to be significant at the 1 percent level (ARI, 1976). The ratio of WCIs for Team A to Team B was 1.03 for week 1, and 0.76 for week 3. That is, Team A, which trained steadily for the entire three weeks, fought progressively better and suffered fewer casualties than the successive Team B's, which were new teams each week and trained only for a week, each. The improvement of the experienced Team A over the fresh Team B's represents a change from an approximately even overall loss exchange ratio (LER) in both teams' first week to a 35 percent improvement in overall LER for Team A by the 3rd week.[4] This may be taken to be an effect of training with time.

Table 2. Average WCI for all Exercises by Weeks

Average Weighted Casualty Index (WCI) For All Exercises by Weeks			
	Week 1	Week 2	Week 3
A Teams	164	126	119
B Teams	159	163	156

The picture changes when casualties are disaggregated according to vehicle and scenario. This was done for tanks, in a search for results more directly comparable with those from the SIMNET IVIS tests. Table 3 (also taken directly from ARI, 1976) shows tank casualties by week for Team A and Team B, separately for the teams in offense and defense. In this case, the LER for Team A in the attack remains approximately the same in week 3 as in week 1 (loss ratio, A/B, of 1.86 in week 1 and 1.78 in week 3); although Team A's absolute tank casualties go down, their kills go down as well. In defense, however, if the data in Table 3 are taken at face value, Team A appears to suffer an 81 percent degradation of LER in week 3 compared with week 1 (loss ratio, A/B for Team A on defense, 0.32 in week 1 and 0.58 in week 3). In almost all cases, the team in defense loses less than the team in offense. However, even though Team A, as it gains experience in defense, exacts more casualties from Team B when the latter is attacking, Team A's losses in defense go up relatively faster. That is, the team that is getting the longer and supposedly better training seems to be becoming relatively worse in defense as time goes by, while conventional wisdom has it that there is a defense advantage (which would presumably improve with training) in being able to fight from hidden positions.

4 Loss Exchange Ratio is defined as the number of Red losses per each Blue loss, and is taken here as the reciprocal of WCI_A/WCI_B with Team A being considered "Blue." When Team B is considered "Blue," the LER is the reciprocal of WCI_B/WCI_A.

Table 3. Comparison of Tank Casualties (by Percentage Lost) for Team A and Team B, Attack/Defense

Comparison of Tank Casualties (By Percentage Lost) For Team A and Team B, Attack/Defense				
	Team A (Attacks)	Team B (Defends)	Team B (Attacks)	Team A (Defends)
Week 1	65%	35% (n=4)	47%	15% (n=3)
Week 2	35%	30% (n=4)	40%	10% (n=4)
Week 3	48%	27% (n=6)	60%	35% (n=4)

NOTE: "n" is the number of engagements on which the result is based.

The limited size of the data set must be taken into account in interpreting these data, and the trend is not smooth by any means. However, the result is counterintuitive, and merits further exploration.

Table 4 shows the results of Table 3 aggregated in a different way. In this table, each team's performance on offense and on defense is displayed separately for each of the three weeks involved in the REALTRAIN exercise. The offense part of Table 4a corresponds to the defense part of Table 4b in describing Team A's performance in offense.

It can be seen that on offense the tanks of Team A made progress from Week 1 to Week 2, and their performance was stronger against the new team they faced in the second week. Team A appears to have been particularly strong in defense, and dominated both of the fresh teams it faced during the first two weeks (even when it was fresh itself, in week 1). However, in the third week, even though Team A was by now well experienced in combat and it faced an inexperienced Team B, the Team A tank performance was about the same in offense as its performance had been in its first week, when Team A was inexperienced, while its performance in defense deteriorated seriously. The fresh Team B's performance in offense in the third week, meantime, was about the same as Team A's had been during *its* first week.

Table 4. Tank Combat Performance of Separate Teams in REALTRAIN

(a) Team A is "Blue." LER = Red Killed/Blue Lost

	Attacks		Defends	
	Percent Lost	LER	Percent Lost	LER
Week 1	65	.54	15	3.1
Week 2	35	.86	10	4.0
Week 3	48	.56	35	1.7

(b) Team B is "Blue." LER = Red Killed/Blue Lost

	Attacks		Defends	
	Percent Lost	LER	Percent Lost	LER
Week 1	47	.32	35	1.85
Week 2	40	.25	30	1.16
Week 3	60	.59	27	1.79

It might be speculated from the results in Table 4 on tank casualties in the successive battles that with the small samples involved the quality of the tank crews in the various teams seriously affects the results and masks any time-based learning effects of Team A armor. Team A tanks may have improved in offense and were very strong in defense during their first two weeks of encounters, and they improved relative to the fresh teams during that period. However, they met a tank force as part of the third-week Team B that was particularly strong--strong enough to overcome the prior two weeks of Team A training and to hold its own against the more experienced tankers.

If there is any validity to these speculations, the only way to reconcile the differing separate results for tanks in the offense and defense scenarios with the overall trend in Team A's favor based on WCI is to suppose that the favorable trend in overall WCI for Team A as training progressed comes from the combined arms aspects of the exercise, and particularly from the meeting engagements that made up approximately half of the total test series. This was indeed the case--*the tank LER in meeting engagements shifted smoothly from approximately even in week 1 to approximately 2 in Team A's favor during week 3* (ARI, 1976, Table 6). Thus, it appears that the effects of several variables are confounded in the tank results, and those results must be considered equivocal at best.

To obtain a cost estimate for the REALTRAIN exercise, the participants in REALTRAIN were estimated from the report to include 36 vehicles and 240 people, or approximately 4-1/2 percent of the size of the average NTC exercise. By simple proportionality, as described earlier, the REALTRAIN exercise was thus estimated to cost $225,000 in FY1989 dollars.

2. Intervehicular Information System (IVIS) Tests

In the IVIS tests one team, the Blue team, fought against semiautomated Red forces for the entire three-week period of the tests. That period included two familiarization periods: one with the M-1[5] tank and one with the M-1 having simulated Block II equipment; a brief base case test period, in which the Blue force performance with the basic tank was measured; and the main test period in which the Blue force performance with the augmented tank was measured. There were 25 test runs, 18 of which provided data describing the runs with the Block II equipment..

Extensive data were taken during all the tests, describing a continuous record with time of tank position, shots, hits and kills for each tank in a shooter-target scoreboard in each test (BBN, 1989). The measures of effectiveness that proved of interest in the present analysis were the range at which each tank opened fire; the numbers of shots, hits and kills for each Red and Blue tank (i.e., the same scores were kept for the semiautomated Red force as for the manned Blue force); and the delay time, t_s, between the first target-shooter intervisibility in any pair (which could be partial) and the first Blue shot in that pair. The time, t_s, was taken as a measure related to, but not to be substituted for, crew reaction time, since the latter could not be measured.[6]

---

5 The test report (HQ, DA, 1988) does not indicate whether the tank was originally in the M-1A1 configuration, but that is not important to this discussion.

6 To ease the computation load involved in matching all pairs several times per minute, this time was calculated for each engagement in which there was a shot, by scanning in 5-second increments through a time "window" around the shot--two minutes before and 1/2 minute after, or until the engagement was over if there was a kill. Events in which there was intervisibility but neither side fired a shot would have been missed in this procedure.

Figures 1a, b and c show the aggregated output of the IVIS tests in the three variables of interest (range at opening shot; t_s ; and losses per kill), for the overall test series, plotted against day in the exercise set. (In all cases of IVIS data, each point in a data plot represents the average output in that variable for all the tanks in the platoon in one engagement.) No strong trend with time emerges, except that the scatter in the first two variables increases markedly in about the second half of the exercise period. The loss data (Fig. 1c) are so widely scattered as to discourage the attempt to fit a curve to them; it would have little meaning.

There is not enough information associated with the data set to suggest reasons for the widening of the data uncertainty during an interval when it might be expected to narrow. Possibly, we are seeing the difference between the early familiarization and base case runs with one version of the tank, and the later runs with the Block-II equipped tank, where something about the tank, the scenarios or the nature of the later trials led to wider variation *with* the added tank equipment than without it. The effects of crews learning how to use new sets of equipment may well be imbedded in these results. But any statements about the causes of the observed trend would be pure speculation in this case. No distinctions among the familiarization runs, the base case, and the Block II cases appear in the data aggregated in this form.[7]

This lack of clearly discernible trends in the overall data progression suggested that the results next be disaggregated, and led to exploration of trends in data representing only the progression with time of the tank with the Block II equipment, and to separation of the data for offensive and defensive scenarios. (And, it must be remembered that the IVIS platoons operated against semiautomated Red forces, so that in reviewing the data with the manned simulator outputs as the basis we are reviewing the Blue performance, only, not the performance of single teams that alternated in the Red and Blue roles. Since Blue played both offense and defense, the distinction in comparison with REALTRAIN is not so great as might first appear.)

---

7 This does not mean that Army analysts learned nothing from the tests about troop performance with the Block II equipment, which was the purpose of the tests. The data in that area were based on C^2 events and sought, in partly quantitative, partly qualitative ways, to describe how the troops used the equipment and, from that, to evaluate whether the equipment would be helpful in combat operations. The Army report indicates general satisfaction with the outcome of the IVIS tests for their purposes. No inference should be drawn from the present analysis of different elements of the same data, for a different purpose, about the value of the tests for their intended purposes.

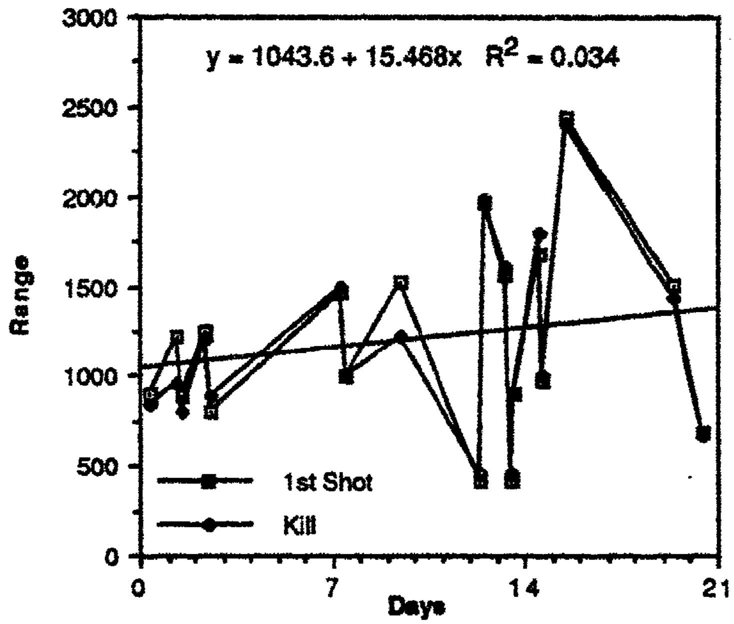


Figure 1a. Range, First Shot and First Kill

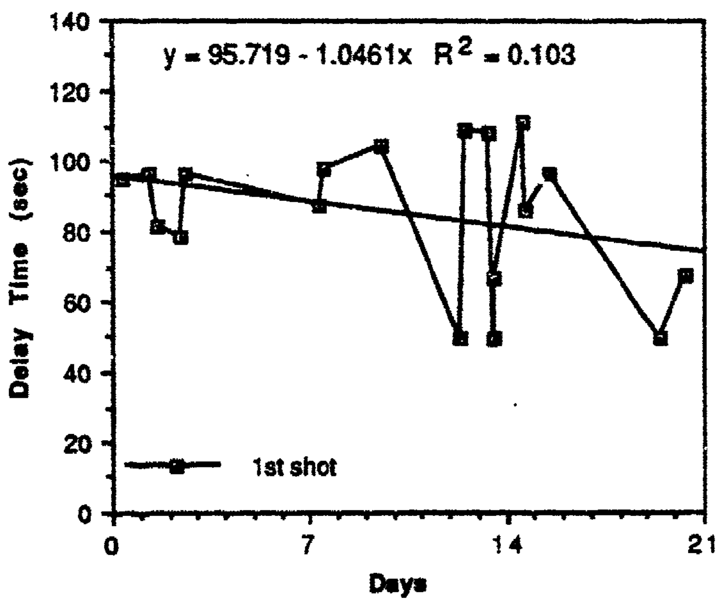


Figure 1b. Delay Time Between First Intervisibility and First Shot

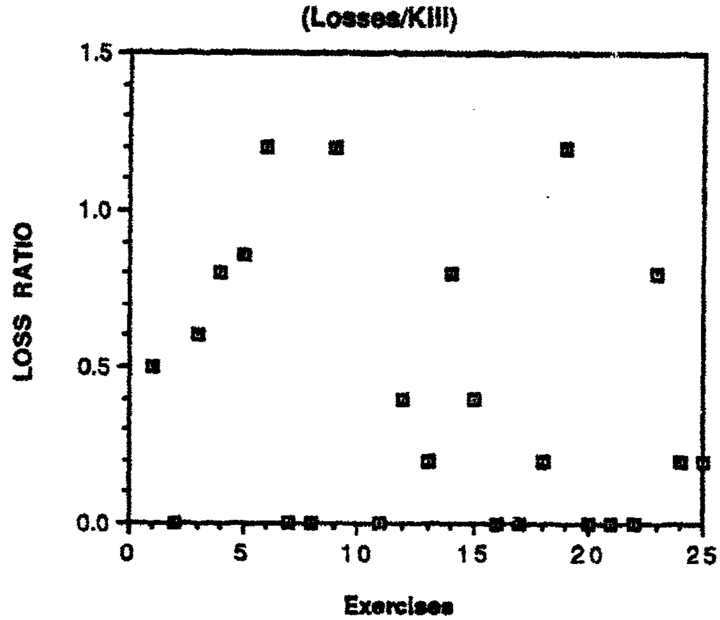


Figure 1c. Loss Ratio

Figure 2 shows the opening range for the Block II equipped platoon on offense and defense; Fig. 3 shows t_S for the same conditions; and Fig. 4 shows the corresponding loss ratio, all plotted against time. Here, some definite trends appear, although the sample size is so small that the magnitudes and (in one case that will be pointed out) the directions of the trends should be accepted with caution until much more experimental data are available to confirm or modify them. It must also be recognized that, since the manned simulator platoon had previous experience in the simulators, leading up to the Block II tests, the trends being observed may be those induced by learning with the Block II equipment, rather than by training in armored engagements per se. This would be consistent with the change in scatter observed for the overall test, discussed above.

Nevertheless, the trends observed are of sufficient interest that it seems reasonable to discuss them as providing at least some provisional insights into training effects in small-unit tank actions, and in the field exercise-simulator network comparison.

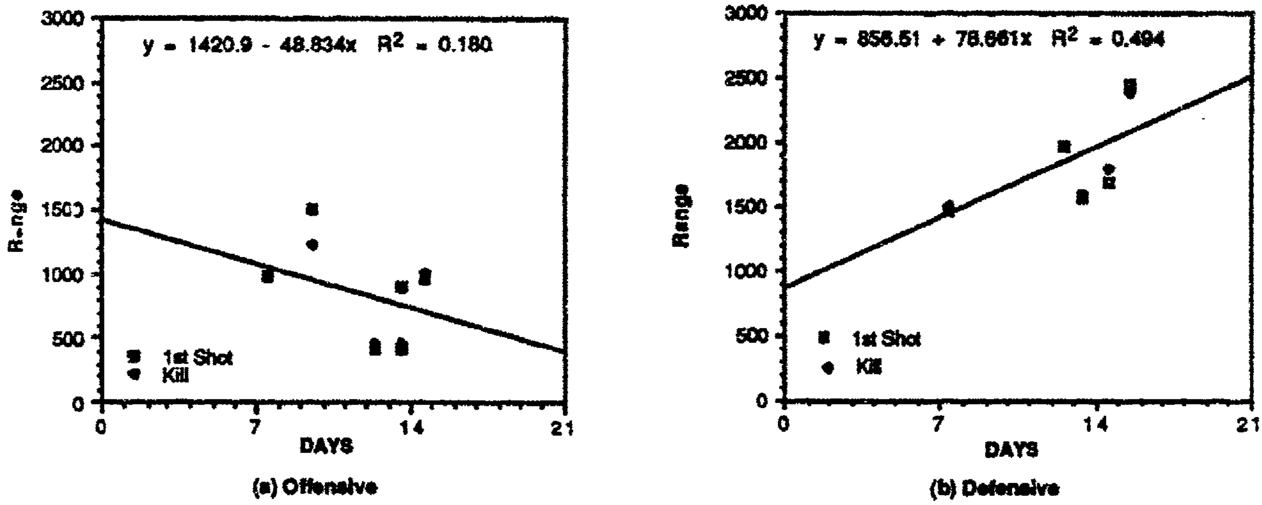


Figure 2. Range at First Shot and First Kill

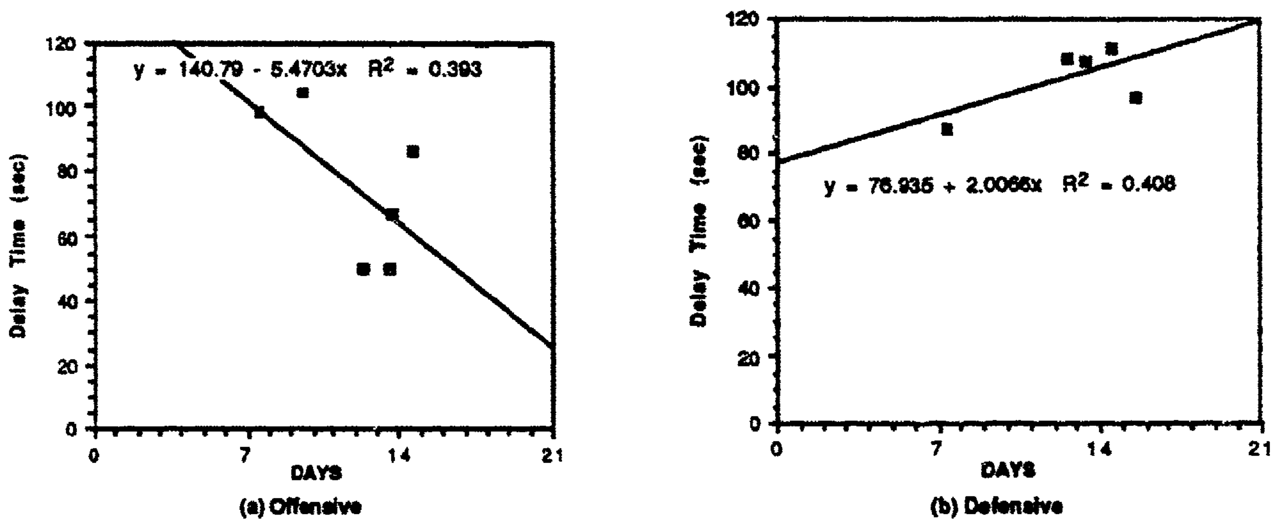
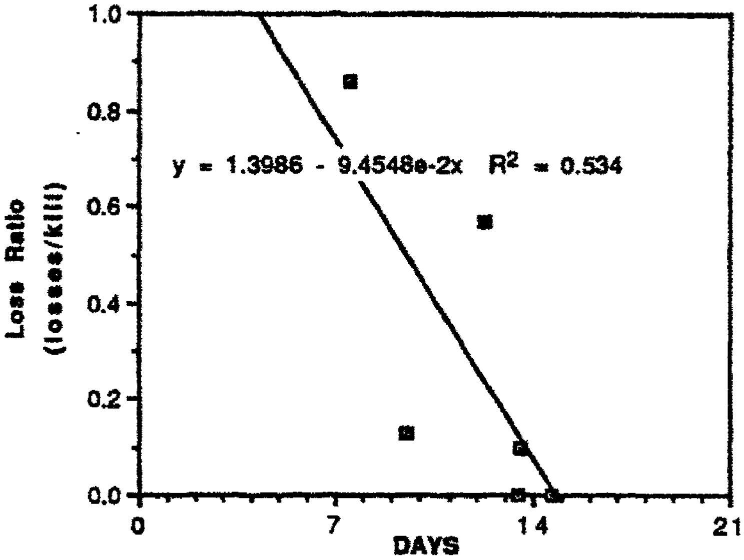
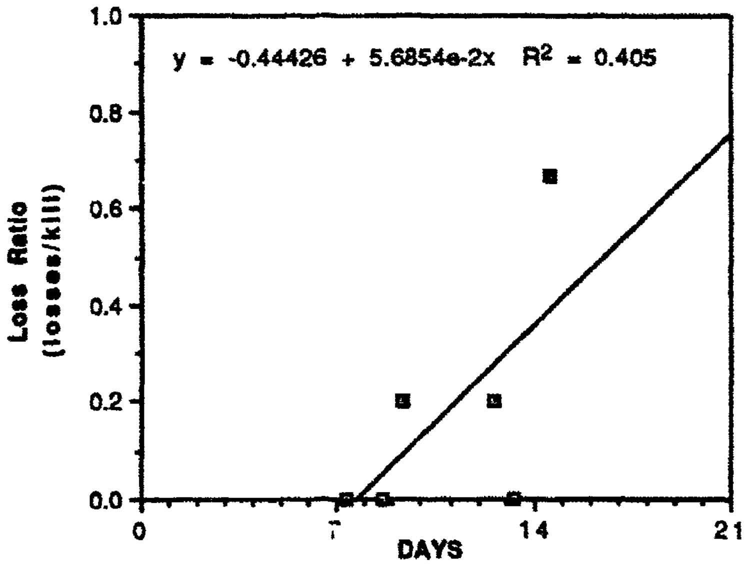


Figure 3. Time Between First Intervisibility and First Shot



(a) Offensive



(b) Defensive

Figure 4. Loss Ratio (Losses per Kill)

Figure 2 shows that the tank platoon on the attack tends increasingly to hold its fire and to open fire on the defenders at closer range as it gains experience. The defenders, however, tend to open fire at longer ranges as they gain experience. The delay time data of Fig. 3 (t_s) complement the range data. The platoon on the attack may hold its fire longer, but the time from first target-shooter intervisibility to the first Blue shot decreases markedly with experience. On the other hand, the time increases somewhat with experience for the platoon on the defensive.

These results appear anomalous. Assuming roughly constant closing speed, decreasing t_s implies earlier target recognition, and if the first shot comes sooner the range at the first shot would be expected to increase. Instead, it decreases. Similarly, if t_s increases with training on the defense, this would be expected to go with a reduction of opening range if the offense's closing rate is constant. Instead, the opening range increases.

It must be remembered, however, that, unlike the REALTRAIN tests, this was not a two-sided test. The same platoon was operating in both attack and defense against a semiautomated Red force. Therefore, the attacker could control his closing rate on attack, and the defender could control his firing delay but not the attacker's closing rate, on defense. The apparently anomalous results can be explained if it is postulated that when on the attack the platoon closed faster with its opponent as it gained confidence in its tactics and knowledge of the terrain. If it closed faster the range at first shot could be shorter even though less time elapsed between first intervisibility and first shot. Similarly, on defense the delay time between first intervisibility and first shot could increase at the same time that opening range to the first shot increased if target recognition came earlier with training. If recognition came earlier the opening range could increase even if the defenders delayed their first shot.

Overall, these results suggest that both sides (or, rather, the platoon in manned simulators in both the offensive and defensive roles) gain confidence as experience with the equipment increases. When on the attack, they move faster but hold their fire until they get closer to the enemy; they then open fire with less delay when the the target comes into view (even recognizing that t_s is not the same as reaction time). When on defense, they hold their fire somewhat longer after the targets come into view, but are able to open fire at greater range because they recognize the targets sooner. These contrary trends in opening range and delay time may indicate that recognition time in terrain becoming increasingly familiar due to repetition, and consciousness of more global relative position advantages than just range to target, could be hidden variables affecting the results.

In the case of the IVIS tests, there was no way to test the hypotheses about why the results appeared as they did, because the tests had been completed long before the analysis described here, and more detailed data about the training effects or the crew reactions within the combat scenario had simply not been gathered. This is an argument for conducting experiments specifically designed to examine the training effects, if possible and affordable, rather than making do with data obtained for other purposes.

The data of Fig. 4, on loss ratio,[8] show a phenomenon that may be related to the tank results of REALTRAIN, in that there were no "clean" and expected trends attributable to learning as training proceeded: the losses per kill go down with experience when the platoon is on the offense, but they increase when the platoon is on the defense. (Intermediate data, not shown for either case, indicate that the number of shots per engagement does not change much during the exercise; the number of shots is related to the rate at which the semiautomated Red forces present themselves and fire, and is therefore not a useful measure of skill growth with experience for these tests. Also, the number of kills per shot are either constant with time or do not show significant variation that would affect interpretation of results based on the losses per kill variable.)

Comparison of Figs. 4a (Blue on the offensive) and 4b (Blue on the defensive) shows that initially the platoon on the defensive has much smaller losses than the platoon on the offensive, and that the latter approaches the performance of the defensive force as experience is gained. The trend toward greater losses per kill by the defensive platoon is driven by a single data point showing high losses in one engagement that appears late in the series; the events surrounding that event are not known, but if its outcome had been in a range closer to the others then the trend in loss ratio for the platoon on the defensive would be essentially level.

Thus, it is fair to observe from this limited set of data that the platoon when on the offensive improves its performance markedly in terms of responsiveness in the battle situation and increased loss exchange ratio, while when on the defensive it appears to improve its operational acuity in terms of engagement dynamics but it holds its own at best

---

8 Note that "loss ratio," used for convenience here, is the reciprocal of "loss exchange ratio" as used in the REALTRAIN results. That is, loss ratio is losses/kill, while loss exchange ratio is kills/loss, both for the Blue side.

and it may lose ground in terms of loss exchange ratio. The platoon on the defensive initially suffers fewer losses per kill than the platoon on the offensive, but the latter approaches the low loss per kill level of the former as it gains experience.

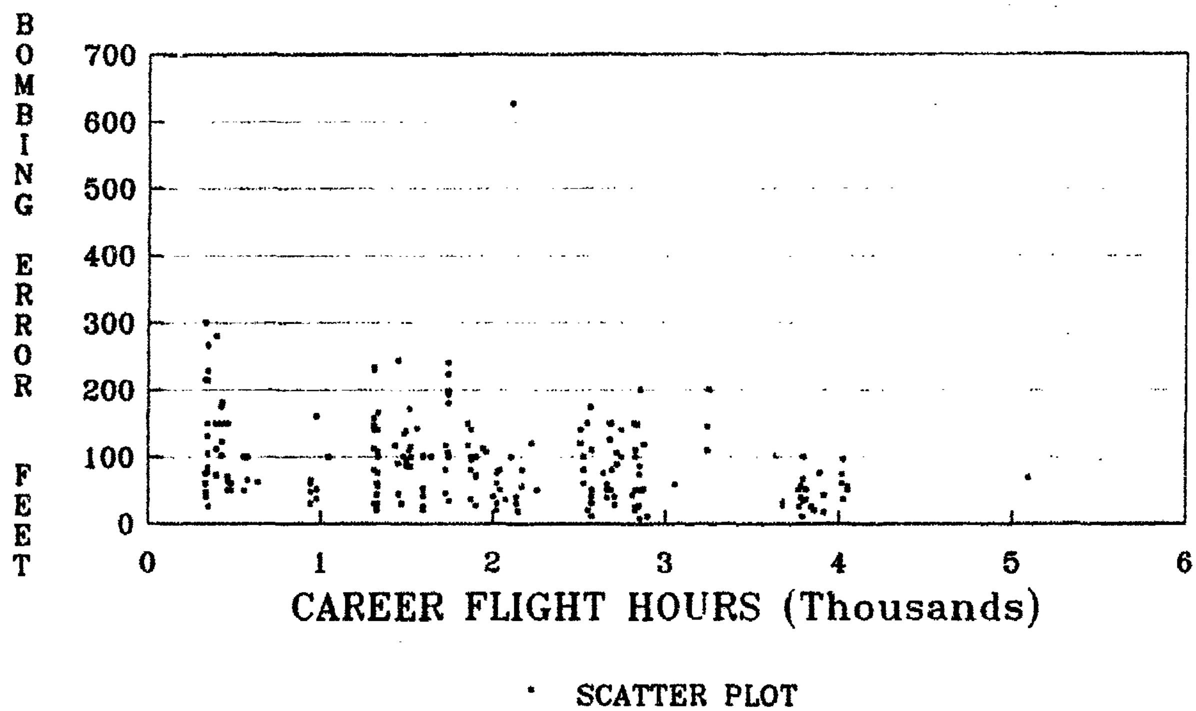
That the platoon on the defensive failed to improve with time, perhaps even more than the platoon on the offensive in view of the "defense advantage" of being hidden and having a better opportunity for the first shot, seems counterintuitive, as was noted for the qualitatively similar REALTRAIN results. Possibly, since the platoon improves in operational acuity while both on the offensive and the defensive, but more so in the former than the latter situation, the initially low casualty level on defense allows less room for improvement, while the initially much more dangerous offensive operation allows for greater improvement as the offensive force learns to find the enemy more quickly and to respond more quickly in ways that enhance its survival and its kill capability.

This explanation is consistent with the data shown, but it must be noted that there is no way, in the restricted data set available, to separate any effects of learning in the use of the Block II equipment that may have been going on at the same time as learning how to fight with the particular tank. The fact that the scatter in the last half of the data sets in Figs. 1 and 2 does not decrease with experience suggests that confounding of combat and equipment learning effects may be small. Alternatively, the expected period of trial and error in learning how to use the new equipment may simply not have been fully traversed during the duration of this exercise.

The costs of the IVIS tests were estimated as described in Chapter II, section G. From the NTC data, assuming 30 military people assigned to IVIS over a 3-1/2 week period, the Service personnel costs come to ~$7000 for the period, compared with the estimated $225,000 for the REALTRAIN exercises. (It should be noted, again, that facilities costs are not included on these estimates. It is not known how the field test/simulator cost relationship might change if they were. However, the training costs might still be expected to be lower, since a large (but unknown) fraction of the field test costs are ascribed to armored vehicle and helicopter movement, not needed for tests with SIMNET.)

3. A-7 Bombing Accuracy

Scores on bombing accuracy are highly scattered (see, e.g., Fig. 5), and only the results of statistical analyses (as distinct from simple observation of regression lines or curves fitted through the data) can say much about underlying trends.



MANUAL AND AUTOMATIC APPROACHES

Figure 5. Miss Distance versus Career Flying Hours

The A-7 results (Mairs, et al., 1986) are presented mainly in terms of the output parameters of regression equations or log transformed histograms and scatter diagrams of miss distance versus diverse experience variables, as illustrated in Fig. 6, which is reproduced directly from the reference.[9]

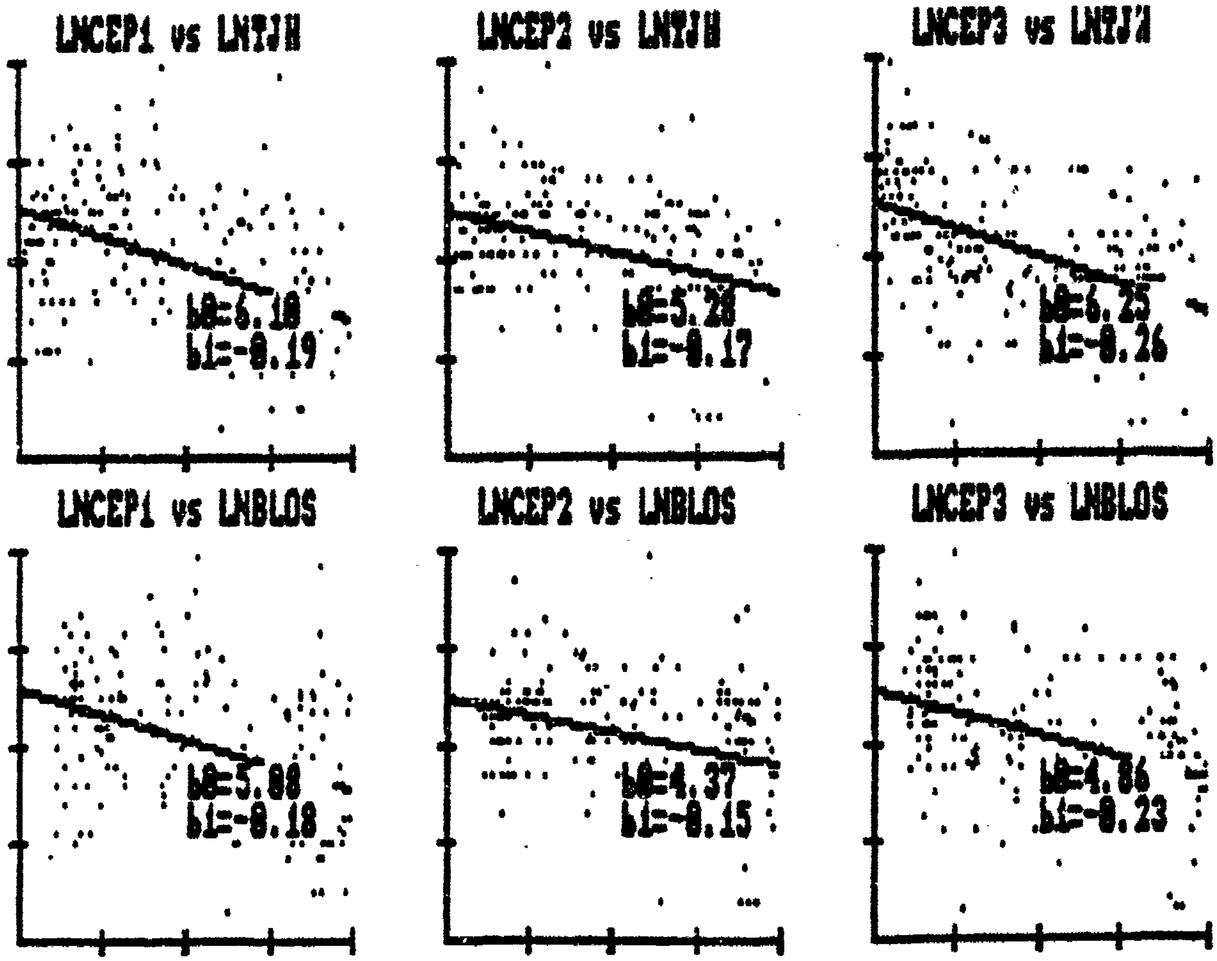


Figure 6. Actual versus Predicted Bombing Scores

(Each dot results from one bombing score)

---

9 Note that while the curves are labeled in terms of "CEP" for the different attack modes, the text does not clearly define "miss distance" as CEP.

Four kinds of bombing are displayed in the data: day/visual in the manual mode; day/visual with computed impact point; night bombing with computed impact point; and laydown bombing. By inference from the reference text, the first three are dive bombing and the last is level bombing. Several experience variables were considered, such as Bombing Length of Service, Months of Flying, Total Hours, Total Jet Hours, etc.

The detailed outputs from this analysis are complex and are used, in the source reference, to discuss issues of readiness, of which bombing scores are one indicator. The "bottom line" of all the comparisons among explanatory variables is that length of service or career jet hours explain equally well the improvement in bombing scores; there is little correlation between individual pilot performance in one kind of bombing and another, of the four examined. Inferences about A-7 training effects for comparison with other data or with equipment improvements can best be made from the simple statement in the reference report that the bombing accuracy has an elasticity of −0.2--that is, a 50 percent increase in lifetime flying hours leads to a 10 percent decrease in miss distance. This includes all the bombing modes, and does not differentiate performance at the beginning of the career experience (~300 hours) and at the end of the experience spread (~2000-3000 hrs); some of the data presented suggest that the point in the career at which elasticity is measured may change the elasticity.

Unfortunately, the A-7 data are not presented in a form permitting easy comparison with the A-10 and F-16 data presented next. The "elasticity" figure was used in the present paper to support the later comparison of A-7 improvement due to training with improvement due to a shift from the A-7 to the F/A-18 bombing system. It was estimated, for example, that a 25 percent improvement in bombing accuracy under certain conditions would require about 375 additional career flying hours, or an improvement of 75 percent in accuracy would require about 1125 additional career flying hours, both starting from 300 hours. The cost of the additional career flying hours (a one-time rather than a recurring cost) was estimated as $9 million or $27 million per squadron in the respective cases, based on $1000 variable cost per flying hour for the A-7 aircraft. This cost could then be compared with the cost of the bombing system change, as explained in Chapter II, section G. These comparisons will be presented in more detail later.

It should be noted in this context that a starting point of 300 hours is low for average pilots in a squadron. According to Horowitz (private communication) the average number of career hours of pilots in a squadron tends to be about 1500, with a median closer to 1000 hours. The increment of flying hours to achieve the improved bombing skill would, by the method used here, then become unreasonably high, as would the cost. This emphasizes the weakness of assuming constant elasticity of accuracy with career hours throughout the career-hour range, and it also emphasizes the degree of uncertainty in any of the cost comparisons involving the A-7 aircraft.

4. A-10 and F-16 Bombing Accuracy

The data presented by Cedel and Fuchs (1986) allow more direct visualization of the relationship between bombing accuracy improvement due to training and that due to equipment change. Figure 7 (reproduced from Cedel and Fuchs) shows the variation of bombing accuracy with mission time (hours of experience in fighters) for pilots in an A-10 and those in an F-16 squadron (and it also illustrates the extent of scatter in the data). For the moment, the progression with pilot experience is of interest, and it is seen to be about the same for the two squadrons.

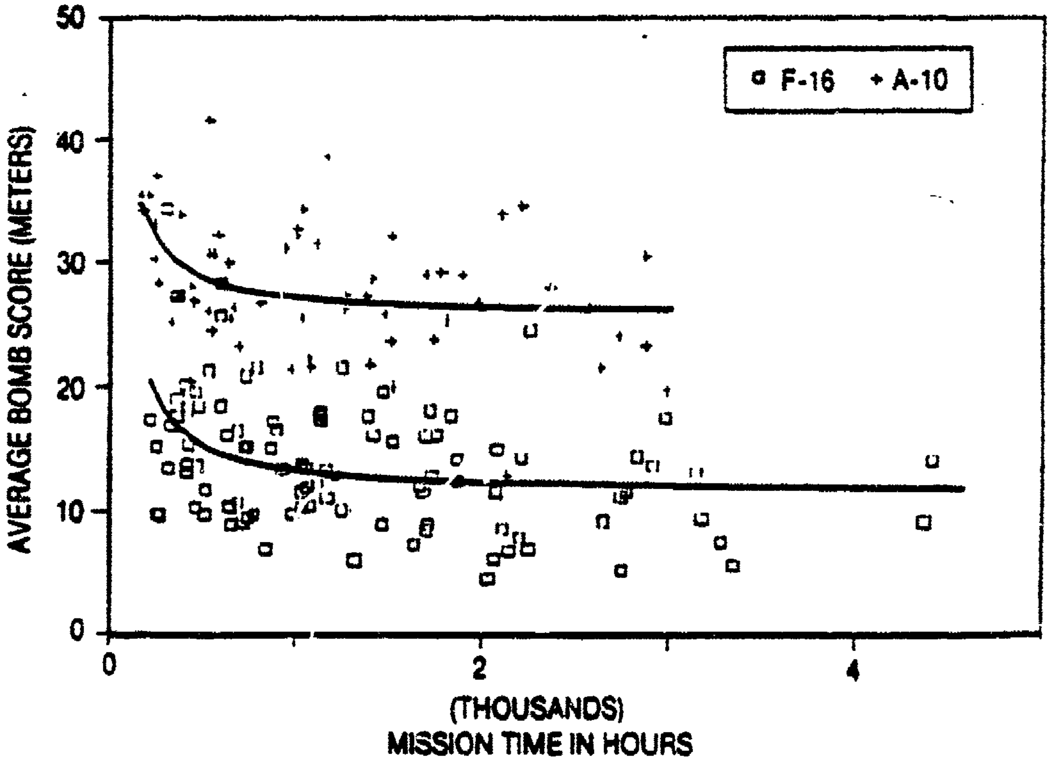


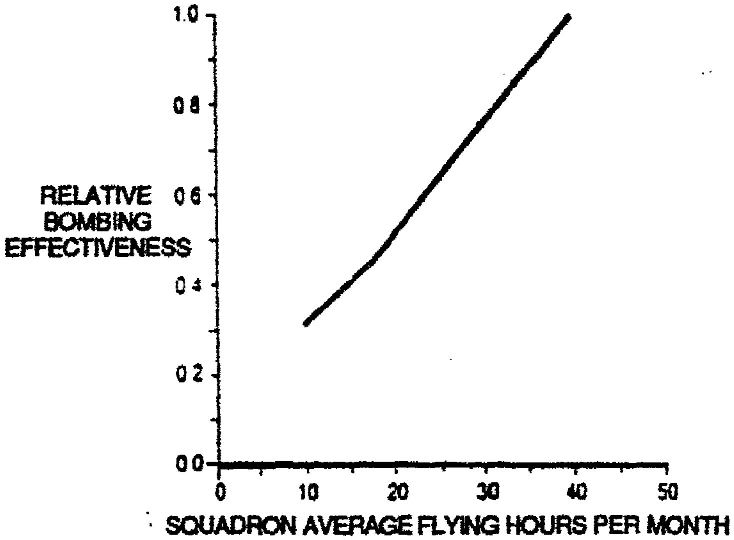
Figure 7. Mission Time versus Bombing Accuracy

Figure 8 (reproduced from the same source) shows the predicted variation of performance with flying hours per month for a squadron using each aircraft, based on a year's flying data. The curves were obtained from a model of pilot performance, developed from the data and showing the loss and gain in proficiency as the squadrons average between 10 and 40 hours per month of flying that includes regular bombing practice.

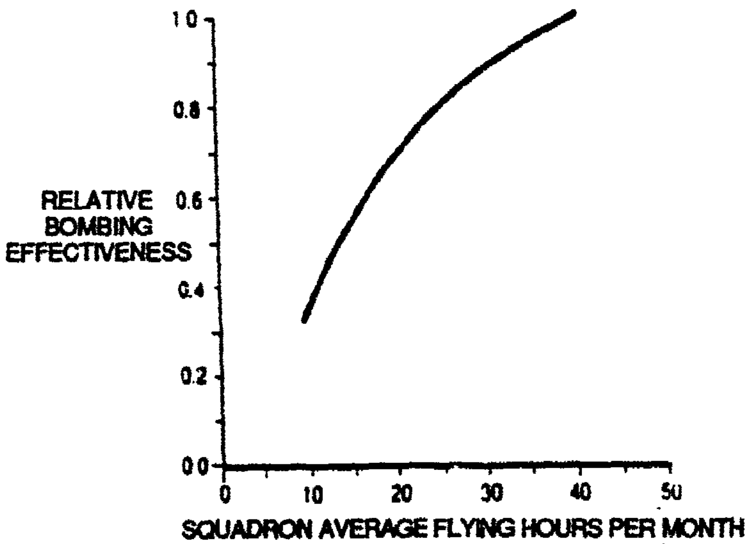
Although the curves of Fig. 8 for the two aircraft differ somewhat in shape, their average slopes and the magnitudes of the changes indicated with practice are roughly the same. The model and data in Cedel and Fuchs are shown by Hammon and Horowitz (1989) to imply about a 60 percent improvement in bombing accuracy from career flight hours, and about 40 percent from recent practice--flight hours per week--for pilots using the two aircraft. Cedel and Fuchs show, further, that there is a career experience threshold--1400 hours for the A-10 and 900 hours for the F-16--below which pilots do not significantly improve their bombing accuracy with immediate practice. Above the threshold, the immediate practice does have an effect on bombing accuracy. From Figs. 5-7, it seems clear that the 60 percent improvement due to career hours comes early in a pilot's career; the improvement at that stage could be confounded with general aircraft familiarization, as well.

The "relative bombing effectiveness" measure in the ordinate of the curves of Fig. 8 is defined as being inversely proportional to the square of miss distance (or CEP). The accuracy measure at 10 hours would thus be considerably less than that at 40 hours, which, from the figures, is the base of reference. From these curves and the defined relationship it was estimated that an increase of flying hours from 10 to 40 hours per month for a squadron of either aircraft would reduce CEP (i.e., improve accuracy) by a factor of 1.8. The cost to do this, for a squadron of 24 aircraft, based on flying hours alone (and assuming the total A-10 cost per flying hour is the same as that of the A-7) is \$0.65 billion over a 15-year period. The 15-year period is taken as the same period that would be used in comparison of life-cycle costs of the two aircraft if the force were upgraded from the A-10 to the F-16 (discussed below).

The results of Mairs, et al. for the A-7 and Cedel and Fuchs for the A-10 and the F-16 are seen to be inconsistent with each other, in magnitude and in details, although both data sets highlight the importance of career flying hours as a variable affecting bombing accuracy. For this reason they were not combined in the present analysis, but will be treated separately in exploration of the two research questions posed initially.



(a) A-10



(b) F-16

Figure 8. Squadron Bombing Effectiveness as a Function of Flying Hours

B. HARDWARE

Improvements due to hardware were chosen for comparison with the data on performance improvement due to training partly on the basis of ready availability and partly on the basis of correspondence to the training cases described above.

1. Tanks

The tank platoon training case could be complemented by a 1974 analysis of improvements in main battle tanks (Graves, et al., 1974). This analysis compared improvements in tanks that were directly comparable with those used in the REALTRAIN exercises.[10] The main improvements relevant to this comparison consisted of the addition of main gun stabilization, enabling firing while on the move; the substitution of an improved fire control computer; and the addition of a night sight. It is not known how these improvements in the tank might relate to learning effects as expressed in tactics and operational procedures, but it appears reasonable to make a "zero'th"-order comparison between the improvements in tank platoon performance derived from increased training with the original tank and those derived from the equipment performance enhancement inherent in the improved tank. There would also be training effects with the improved tank; this issue will be discussed in connection with the overall comparison of results.

The MBT analysis showed that the aggregate measure of effectiveness in platoon-on-platoon combat against a constant threat tank, encompassing the LERs for tanks on the offense and the defense in a variety of terrains, improved 20 percent in the upgrade from the M-60A1 to the M-60A3. The cost to achieve this upgrade was projected as 10 percent of the 10-year system cost. Translated into 1989 dollars, the cost of the upgrade for a unit of 5 tanks would be $2.83 million over 10 years.

2. Improvement From A-7 to F/A-18 Bombing Systems

The hardware part of the training/hardware-improvement comparison involving the A-7E and F/A-18 aircraft was based on the spread of differences between the two aircraft in day-visual bombing and in radar bombing. The main training parameter used was the –0.2

---

10 Although the REALTRAIN report does not identify the tanks actually used by the participating units in Europe, the elapsed time between the analysis of Graves, et al, and the REALTRAIN tests (Oct., 1974-Nov., 1975) makes it very unlikely that the upgrade from the M-60A1 to the M-60A3 reached operational units in the time interval.

elasticity of accuracy with career hours given in Mairs, et al (1986), as indicated above (section A3), which was said to apply to all conditions. The Joint Munitions Effectiveness Manual shows an overall spread of bombing accuracy differences between the two aircraft of from about 15 percent to about 75 percent, depending on the bombing conditions. This spread results mainly from differences in the bombing radar and the bombing computer; the F/A-18 bombing system also has night attack equipment, but that did not figure in the comparison being made.

The cost difference of the radar and bombing computers in the two aircraft (including the fuze function control set) comes to $690,000 per aircraft, or ~$17 million per squadron. This way of looking at the cost implies that the new bombing system is simply fitted to the A-7 aircraft. The comparison between training and hardware improvement in this case would thus assume that this retrofit could be made and that all other things about the aircraft were equal--that is, that the flying qualities, cockpit layout, etc., do not affect bombing accuracy.

Another way of looking at the hardware cost would be to consider the difference in total cost of the two aircraft. This would be more consistent with the A-10/F-16 comparison to be made. The difference in 15-year system costs between a squadron of A-7s and F/A-18s is $480 million.

3. A-10/F-16 Comparison

There are many differences in mission between the A-10 and the F-16 aircraft, since the F-16 is a fighter and an attack bomber while the A-10 is a bomber only. However, the two are treated here as though all the cost differential between the aircraft should be attributed to the bombing mission. This implies more cost for the F-16 as a bomber than is warranted, since the F-16 can also be used as an interceptor and an air superiority fighter, while the A-10 cannot.

The curves fitted to the data in Fig. 7, section A4, show that the F-16 is consistently about twice as accurate a bombing aircraft as the A-10, through the entire range of pilot mission hours. The difference in 15-year program cost between an A-10 and an F-16 squadron is $600 million; this is considered to be the hardware cost of achieving about a factor of 2 improvement in bombing performance, for comparison with the cost of achieving about the same result by enhanced training of A-10 pilots. This neglects for the time being the fact that, according to Cedel and Fuchs, the same training achievement could be obtained for the F-16, and it also neglects the important caveat about aircraft mission differences noted in the previous paragraph.

4. Automated Equipment

While not in the "mainstream" of this investigation, it is worth taking note of some data that emerged regarding the interaction between automatic equipment and training. At some point during future explorations of the military value of training it will become important to consider those interactions (as will be discussed in the next section), and the data fragments discovered carry implications, as yet weak but worth exploring, about the influence of equipment design on the extent of training needed to achieve improved unit skill levels.

The tank data were obtained in an evaluation of the performance in M-60 and M-1 tanks of 1131 7th Army tank crews falling into different mental categories on the Armed Forces Qualification Test (AFQT) (United States Military Academy, 1984). The performance measure was equivalent tank kills, which was estimated from numerical scoring of a series of main gun and machine gun firings by tank crews at the Grafenwoehr, West Germany test range over the period Jan-June, 1984. Only the tank commander (TC) and the gunner were included in the rating measure.

The key results of interest here, shown in Table 5, were that while crews of both tanks improved in performance as their AFQT categories increased over the Category IV base case, the improvement was much smaller in the M-1 than in the M-60 tank because the baseline performance in the M-1 was much higher to begin with, as shown in Table 6. The implication is that sophisticated equipment with a higher degree of automation, which characterizes the M-1 relative to the M-60, substitutes for a lot of crew capability.

Similar results, more directly applicable to training, have been obtained for the F/A-18 and the AV-8B in bombing. Figure 9 shows the expected values of bombing accuracy with career flight hours for the two aircraft in the manual and the automatic modes, obtained from data such as those illustrated in Fig. 5. Unlike the bombing accuracy improvements with career or mission flight hours obtained in the manual bombing modes for all the aircraft examined, the two aircraft in the automatic mode show the same level of accuracy across the entire illustrated range of career flight hours, and the accuracy

Table 5. Percent Increase in Performance
M60
M1

TC Mental Category					
Gunner Metal Category	I	II	IIIA	IIIB	IV
I	75.17% 18.9%				
II		62.84% 16.3%			
IIIA			45.9% 12.4%		
IIIB				27.9% 8.0%	
IV					Base Case

Table 6. Percent Increase in Performance Due to the M1

Crew Mental Category					
	I	II	IIIA	IIIB	IV
Percent Improvement in Crew Performance M1 Over M60	+25%	+31%	+41.43%	+55%	+84%

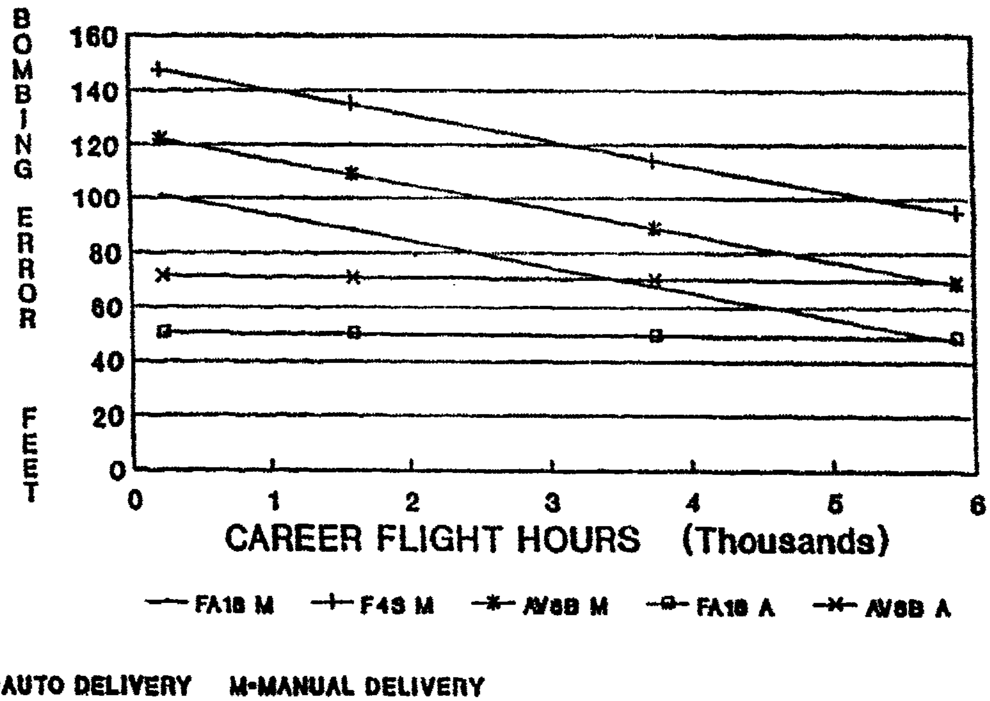


Figure 9. Bombing Errors versus Career Flight Hours Previous 7 Days at Mean

level is comparable to the best achievable in the manual mode.[11] The implication here, too, is that sophisticated equipment with a high degree of automation substitutes for an extensive amount of training.

These results suggest avenues for further consideration in the broader context of factors affecting the military value of training as compared with the military value of equipment improvement.

---

11 The curves shown in Fig. 9 result from statistical estimation of the parameters of a single equation descriptive of the entire data set (Hammon and Horowitz, 1989). Terms were included in the equation that allow the prediction of different flying-hour effects for manual and automated runs and different levels of accuracy for each type of aircraft in the sample. The form of the curves is thus derived from the assumed function underlying the equation. The curves are the expected values of the performance that emerge from the statistics. Individual regression curves fitted to the bombing performance data of the separate aircraft show different slopes, and sometimes the interaction between flying hours and the presence or absence of automation was not as clear as it appears in Fig. 9. Further analysis of these statistics (such as experimentation with different functions fitted to the data) and the underlying phenomena are needed before the results of Fig. 9 can be accepted as final. Such analysis is under way as part of a different project at IDA (S. Horowitz, private communication).

IV. INTEGRATED COMPARISONS AND DISCUSSION

A. PRESENTATION OF OVERALL COMPARISONS AMONG RESULTS

It is now time to put all the results presented above together in a coherent way, and to evaluate what has been learned from all the fragments of data and information presented thus far. The results of the above separate analyses are combined and compared in different contexts in Tables 7, 8, and 9 for armored combat and tactical air bombing.

1. Tank Combat Comparisons

Table 7 shows the comparison between the REALTRAIN and the SIMNET IVIS results, in terms of the difference in effectiveness outcomes between the two approaches, and the cost of each. The cost figures are given for single exercises in each medium; as will be noted further below, the extent of training, in terms of numbers of exercises needed to obtain the results or to meet the (presumed constant) improved proficiency level due to hardware advancement, is not known. While REALTRAIN indicated a 35 percent improvement in overall exchange ratio, the small sample size and the data scatter for the IVIS tests do not permit any comparable figure to be derived from those tests.

The two approaches yield similar results in one important respect, namely, that the difference in learning effects on the outcomes of battle when Blue is on the offense and when Blue is on the defense are roughly similar. That is, in both cases, the Blue loss ratio is initially significantly higher when Blue is on the offense than when Blue is on the defense. This is not unexpected, since tanks on the offense are exposed and those on the defense are usually hidden. Blue's operations in detail on the defense may improve but progression in their loss ratios is equivocal, while the offense's loss ratio declines and approaches that of the defense as experience is gained.

This similarity in an important trend suggests that while the results obtained in the two media might be found to differ in overall magnitude and in environment-dependent details if there were comparable data, they nevertheless appear to be qualitively comparable, and comparative trends in small unit capability move in similar directions in the two media.

Table 7. Comparison of Field and Simulated Outcomes
Armored Warfare--Platoon-sized Action--Three-Week Period

COMPARISON PARAMETERS		REALTRAIN (Combined arms platoon)		SIMNET, IVIS TESTS (Tank platoon)
EFFECTIVENESS	MOE	Casualty Ratio, B/R, from Wtd. Cas. Index	tk loss ratio B/R, wk 3 / tk loss ratio B/R, wk1	Range opened fire; time, 1st intervis. to 1st shot; loss & kill ratios
	SIZE/DIR'N OF EFFECT	35% improvement in Blue's favor over 3-wk period	Offense: No change; Blue losses & kills both declined 25% Defense: Blue tank improvements inconclusive	- Blue offense fired at shorter range as gained experience: time delay to firing decreased - Blue defense opened fire at longer range as gained experience; time delay increased - Blue offense LER increased; Blue defense LER decreased - Red loss rate initially much lower than Blue's - Inference: Offense became more aggressive and skillful, probably incl. tgt recognition in the terrain; reduced loss rate despite "defense advantage"
	QUALIT-ATIVE EFFECTS	Since Blue LER did not change on offense, & Blue change inconclusive on defense, improvement in overall WCI must come from meeting engagements and combined arms (confirmed by data)		
REMARKS ON DATA		- Data in two cases not directly comparable; sample in each REALTRAIN scenario, small; through SIMNET sample - Meeting engagement in REALTRAIN, most like offense in SIMNET - Combined arms platoons not same, but # tanks are		- Sample too small for change in MOE to yield significant stastistical comparisons - Data not taken for training purposes, but to evaluate equipment
COSTS ($ '89)	COSTS OF WHAT	Three-week, 2-sided exercise by two combined-arms platoons (Tank is M-60)		Three-week SIMNET-D exercise with 5 M-1 manned tank simulators, to test Inter-Vehicular Info. Syst. (IVIS)
	DOLLARS OR RATIO	$ 225, 000; does not include fixed training facility cost (latter imbedded in total Army training budget)		$ 7,000; same basis as REALTRAIN costs
REMARKS ON COSTS		Obtained from ratio of force sizes, multiplied by typical avg. cost for Bde-sized exercise at NTC		Obtained by same method as REALTRAIN costs (NOTE: no vehicle moves involved)
		COSTS DO NOT INCLUDE TROOP PAY & ALLOWANCES IN EITHER CASE		

OVERALL OBSERVATIONS:

1. Qualitatively similar results, field and simulation
2. Samples too small for quantitative similarity to be statistically significant
3. SIMNET costs 1/30 as much as field exercise for similar test, but fixed facility costs not included in either case (not available via standard accounting system)
4. Improvements in capability favor offensive operations; consistent with doctrine of armor as a primarily offensive force
5. Difference in tanks and force composition could affect comparison of results, but similarities in key areas are encouraging

NOTE: For sources, see listing in comprehensive table in Summary

The parts of the costs that could be included in the comparison show the simulation to be much less expensive than field training. The facilities cost comparison between a simulator facility of fixed, roughly battalion size and a large training range with its full complement of facilities and support personnel able to handle brigade-size exercises would seem, intuitively and pending further investigation, to favor the simulator facility.

One would not always do simulator training solely to save costs. However, the cost results, together with the qualitatively similar effectiveness results, do suggest that when the special penalties of the range and the size force that the range can handle are not needed, the simulator facility can be used to provide a good deal of useful, economical, more controllable training activity. This conclusion is reinforced by the relative ease of controlling and replicating critical test parameters in the simulator environment, and the easier measurability of the simulator outputs. In addition, it may take from 12 to 18 months to field prototype equipment needed for field trials, while tests can be performed within 3 to 6 months in the SIMNET, adding time as an additional factor in favor of the simulator if the unique qualities of field tests are not needed. This result is not unlike that obtained for flight training, although the appropriate, relative levels of simulator and actual field training in the two cases doubtless differ.

Table 8 compares the level of improvement and cost in the platoon-size tank battle outcome due to training and due to tank hardware improvement. These comparisons are based on the REALTRAIN exercise and on the change from the M-60A1 to the M-60A3 configuration; thus the starting point, a platoon using the M-60A1 tank, is the same in the two cases (except for differences in platoon composition).

As can be seen, the improvements in unit capability are roughly of the same magnitude; the REALTRAIN exercise leads to an overall improvement across the spectrum of scenarios and combat conditions of about 35 percent in the loss exchange ratio, while the tank improvement leads to an improvement of about 20 percent. Given the scenario, model, data source, and force composition differences, agreement within less than a factor of two must be considered fairly good; certainly, an order of magnitude disagreement would not have been surprising, under the circumstances. (Of course, any agreement between the two methods of force improvement must be taken as an observation, not an expectation. There was no *a priori* reason to expect any agreement at all; the extent of agreement or lack of it was implicit in the research questions to be answered.)

The outcome of the cost comparison depends on how much training is necessary and on the impact of the unknown facilities costs. It appears that if most of the training is done in the field, even without the facilities costs, and if more than one training exercise per year is needed, then the equipment will be less expensive. The ratio would be far different if much of the training could be done with the simulation network, since the cost for training, even with several sessions per year, might then be on the order of a tenth as much as the hardware improvement.

All this begs the question that training would be necessary with the new as well as the old tank. That issue and its significance will be discussed later, together with a similar outcome for the tactical air bombing case.

2. Tactical Air Bombing Accuracy Comparisons

Table 9 shows the results of the comparisons made in the case of tactical bombing accuracy. The results for the A-10/F-16 are more informative because they shed light on all the factors that "play" in the comparison: career flight hours; recurrent training hours; and the improvements due to hardware. Both sets of comparisons are instructive regarding the nature of the comparison problem and the research questions posed initially, however.

The table shows that increasing current flight hours from 10 to 40 per month in either the A-10 or the F-16 aircraft, with some unknown part of that time devoted to bombing practice, leads to just under a factor of 2 improvement of squadron average bombing accuracy, for pilots with over 1400 mission hours. It also shows that if the squadron were to change aircraft, from the A-10 to the F-16, it could achieve about a factor of 2 improvement in overall bombing accuracy for all levels of pilot experience. It is not known from the experimental data whether these two effects are additive--i.e., whether shifting to the new aircraft and increasing current training time would improve bombing accuracy by a factor of 4. The presumption from the data is that it would. But it must be emphasized that these are average results from widely scattered data, so that they would not predict improvement in any single bombing run of the magnitude illustrated.

The cost to achieve the improved average accuracy by training or by going to a better aircraft is about the same--on the order of $600 million over a 15-year period. The training cost is the total flight hour cost of 30 hours per month for a squadron for 15 years, and the reequipment cost is the 15-year life cycle cost of a squadron of F-16s that would replace the A-10s.

Table 8. Comparison of Training and Equipment Outcomes in Armored Warfare Platoon-Sized Action With M-60 Tanks Over a Three-Week Period

COMPARISON PARAMETERS		TRAINING REALTRAIN (Combined arms platoon, M60 tanks, assumed M60-A1 from time period)	EQUIPMENT Tank Engagement Model, comparing M60-A1 with M60-A3 tanks
EFFECTIVENESS	MOE	Weighted Casualty Index Ratio, B/R, in engagements over 3-week period	% improvement, A1 to A3, composite LERs in multiple engagements on diverse terrains
	SIZE/DIR'N OF EFFECT	35% improvement in Blue's overall Loss Exchange Ratio over 3-wk period; results for tank part of force alone equivocal; overall improvement derives from combined-arms meeting engagements	Blue composite Loss Exchange Ratio increased 20% vs. Red in going from M60-A1 to M60-A3 (Tank improvement includes better fire control computer, night sight, shoot while move)
	QUALIT-ATIVE EFFECTS	Blue platoon operated continuously over 3 weeks, while new Red platoons with less trained troops were entered each week	
REMARKS ON DATA AND ANALYSES		- Combat units in two cases not directly comparable; platoons in REALTRAIN include TOW and infantry squads in APCs; platoons in Tank Exchange Model use tanks, only - Though platoons not same, # tanks are - Both sides use same tank in REALTRAIN; equipment analyses use US vs Soviet tank	- Comments in training column apply - Equipment-change effectiveness comparisons are in anticipation of new equipment performance; training effectiveness estimates are from field measurements
COSTS ($ '89)	COSTS OF WHAT	10-year cost of three-week, 2-sided exercise by two combined-arms platoons, 1 time per year or 4 times per year (range cost not included)	10-year life-cycle cost difference between a platoon of M60-A1s and a platoon of M60-A3s
	DOLLARS OR RATIO	$2.25 Million if 1 time per year; $9.0 Million if 4 times per year	$2.8 Million
REMARKS ON COSTS		Obtained from ratio of force sizes, multiplied by typical avg. cost for Bde-sized exercise at NTC COSTS DO NOT INCLUDE TROOP PAY & ALLOWANCES IN EITHER CASE	— — — — — —

OVERALL OBSERVATIONS:

1. Improvement in effectiveness measure due to training: 0.15/million if one training session/yr is sufficient for skill retention; 0.04/million if four sessions/yr are needed.
2. Improvement of equipment effectiveness measure is estimated to be .14/million, before acquisition of new tank.
3. New equipment would require training also.
4. Performance improvement with new equipment unknown.
5. Degree of training needed in both cases unknown.
6. Total improvement with new equipment would be function of equipment and training improvements; degree to which additive, unknown.
7. But training with old equipment could give roughly similar results to equipment improvement, until outclassed by enemy equipment improvements.
8. CONCLUSION: Equipment/training trade-offs appear feasible but must be done judiciously; insufficient data available as yet to decide how

NOTE: For sources, see listing in comprehensive table in Summary.

Table 9. TACAIR Training/Equipment Comparisons

COMPARISONS MADE and MOE	RESULTS: TRAINING	RESULTS: EQUIPMENT	REMARKS
<u>A-10 with F-16</u> Change in bombing CEP with Flight Hours	- Squadron average CEP @ 10 hrs/mo is 1.8 X squadron average CEP @ 40 hrs/mo flight time - Results approx. the same for A-10 and F-16		- These are statistical results; underlying data highly scattered - Absolute avg. acc. different for pilots >or< 1400 career hrs (1) - Other data analyses show that ~60% of training effects come from career flight hours, ~40% from recent practice (2)
Change in CEP with A/C type		A-10 average CEP is ~2X F-16 average CEP, regardless of pilot career experience	Based on <u>measured</u> CEP, same as flight hour data for training (1)
Cost to achieve result ($'89, 15-year cost, for flight hours or difference in A/C, respectively)	$650 million for squadron flight hours	$600 million more for F-16 than for A-10 squadron	Training costs based on cost /flight hour; do not include pay & allowances; assumes A-10 costs ≈ those for A-7
<u>A-7 with F/A-18</u> Improved CEP in bombing	A-7 CEP reduction of 25% would need ~375 career flight hours, for relatively inexperienced pilot (300 hrs to start); 75% reduction would need 1125 hrs.	15% to 75% reduction in CEP in going from A-7 to F/A -18, depending on bombing conditions(3)	Bombing improvement with training inferred from "elasticity" of -0.2 in ref.(4) independence from career flt hrs, questionable (cf, data in 1)
Cost to achieve result (Same basis as for A-10/ F/16, above)	$9 million/sqdn for 25%; $27 Million for 75%. One-time cost, since based on career hours, not recurring costs -- see remarks)	$17 million/sqdn, procurement (5), if F/A-18 bombing system assumed to fit A-7, and procured as one-time cost for A-7 only; $480 million 15-yr cost increment, if procure squadron of F/A-18s	Even if F/A-18 bombing system fit A-7, A/C flight characteristics could affect bombing accuracy

1. USAF, *An Analysis of Factors Affecting Pilot Proficiency*, Dec., 1986
2. IDA , Hammon & Horowitz, *Flying Hours & Aircrew Performance* (Working Paper), Jun. 1989
3. Inferred from JMEM
4. *Quantifying the Training to Readiness Link: ... Attack Pilots*, SAG Corp., Nov. 1986
5. Equipment only; does not include retrofit costs
6. Cost data for A/C, from SAR, by M. Olver, IDA Cost Analysis & Research Div; all other cost data developed by J. Stahl, IDA CARD; variable flight hour costs for training with retrofit; total flight hour costs for training with new A/C; see cost discussion in text.

The A-7//F/A-18 comparison is less informative by itself because the training data are much more sketchy and the aircraft hardware comparisons are not based on experimental results as are the comparisons for the other two aircraft. However, they allow us to build on the A-10/F-16 results to gain further insights into the nature of the comparisons that must ultimately be made.

According to the available A-7 data, it would cost much more to achieve the maximum improvement in bombing accuracy by increasing career flight hours than it would cost to improve the bombing system. However, if improving the bombing system alone were not enough, because aircraft flying qualities were important and the A-7 could not be retrofitted with the F/A-18 bombing system, so that one would have to change the entire aircraft system (the comparable case to shifting from the A-10 to the F-16), then the hardware/training cost relationship would reverse: it would cost significantly more to achieve the levels of improvement shown by shifting to the more capable aircraft than by training. Note, however, that the training data being compared relate to career hours rather than to current flight hours and bombing practice. If it could be assumed that the improvement with current practice would be about the same for the A/7 as for the A/10, then the total cost to achieve and maintain the 75 percent or so improvement of proficiency with the A-7 by training would become roughly comparable to the cost of going from the A-7 to the F/A-18 aircraft.

Overall, these results suggest (not yet very strongly, because the data are so meager and not presented in comparable form or conditions in the sources) that improving avionics to achieve bombing accuracy improvements may be significantly less expensive than simply doing more training with inferior avionics, but that if the entire aircraft system must be improved by acquisition of a more modern aircraft then the costs of training or of the hardware improvements may be comparable, and more expensive by a wide margin than changing the avionics.

The improvement in capability illustrated by the A-7//F/A-18 hardware comparison is not quite as large, at best, as the improvement shown by the A-10/F-16 comparison. First, the data sources are different, the former being analytical and the latter being experimental; and second, there are no A-7 data showing improvement with current training that can be compared with the A-10 and the F-16.

The large range of possible improvement between the A-7 and the F/A-18 that was obtained analytically is of interest, however. The top end of the range for the F/A-18 shows potential improvement roughly comparable to that for the A-10 and F-16, but at other conditions the potential improvement is much smaller. Such a phenomenon could be acting in the A-10/F-16 case as well; the reference indicates that all the data for the two aircraft were obtained under the same bombing conditions--low-angle dive bombing with low-drag bombs. The A-7//F/A-18 range of results determined analytically cover the entire spectrum of delivery tactics and weapons, from day-visual dive bombing to level radar bombing at night, with conventional, low drag, and retarded bombs. The maximum and minimum improvements do not necessarily correlate with bombing scenarios or conditions, but day-visual bombing (the A-10/F-16 case) requires the most pilot skill and would therefore show the greatest improvement with training. It is possible that over the broad range of tactics and weapons the A-10/F-16 comparison would show a similar range of variability to that in the comparison between the A-7 and the F/A-18 (but note: the A-10 does not have a radar bombing system, although the A-7 does).

B. DISCUSSION OF OVERALL RESULTS

1. Research Question 1: Magnitude of Improvement and Military Value

1. Can realistic, quantitative values for unit training effectiveness be determined that would lend credibility to model-based calculation of the military value of training expenditures?

The results of this analysis show that training yields quantifiable improvements in performance, in the areas of warfare examined. The magnitude is sizeable, especially in the case of tactical bombing, but does not appear to offer a decisive change in the military balance as the latter was assessed in the earlier work (Deitchman, 1988).

The comparison of tank combat capability improvement with either training or with hardware improvement (Table 7) shows that the performance of a platoon-sized tank force can be improved on the order of 20-35 percent. The higher end of the range is the field training result, but the tank platoon in that case was augmented by infantry and anti-tank weapons. Thus, it is not known whether the greater improvement in unit performance in the training case over the hardware improvement case comes from the difference in unit composition ("pure" armor versus combined arms) or from the fact that the analytical model used for the hardware comparison does not capture some unknown but very important human factors effects that are reflected in the training data. It is also not known whether a certain amount of improvement in the performance of all platoons might be magnified into a larger performance increase of companies, battalions or brigades in which the platoons would be imbedded. At any rate, the samples are too small to suggest more than the general order of magnitude of the change in the two cases.

Figure 10 reproduces the results of IDA Paper P-2094 (Deitchman, 1988) that show the impact of improvement in armored force performance on the outcome of the war played in the TACWAR model. It can be seen that an improvement of 20-35 percent would have little impact on the ultimate result. To achieve the necessary factor of 2 or more improvement in armor capability, one would have to shift from the M-60 series to the M-1 series of tanks; other analytic studies show that much improvement or more in moving from the older to the newer tank (a quantitative experimental comparison of unit performance with the two tanks on terms comparable with those in this review has not yet been found, and may not exist).

Training would still be necessary to achieve unit proficiency in fighting with the M-1 tank. The results in the U.S. Military Academy (1984) report suggest that the high level of automation in the M-1 tank might lead to a lower overall requirement for time and effort devoted to proficiency training for individuals, but that would not necessarily extend to the aspects of performance that characterize unit operations on the battlefield.

The most important additional result applicable to the military value of training that this exploration has elicited is the indication that tank platoons on the offense greatly improve their loss exchange ratios as training over a three-week period proceeds, and gradually approach the higher level of exchange ratio achievable from the start by the defense. There is the further indication that this value of unit training can be achieved at significantly less expense by performing most of the training in a simulator network like SIMNET.

Figure 11 reproduces the results of the earlier exploration (Deitchman, 1988) for the case of improvement in tactical air-to-ground warfare. It can be seen that a factor of 2 improvement in bombing effectiveness (shown by Table 9 to be potentially achievable through training *or* hardware improvement) can make a noticeable difference in the outcome, but that to reverse the outcome of the war in the TACWAR model this bombing effectiveness improvement must be accompanied by a factor of two improvement in aircraft survivability. No experimental data were found during this exploration that showed the impact of training on combat survivability in TACAIR (ability to avoid being shot down by air defenses while delivering weapons). However, the circumstances of the data described here are not encouraging.

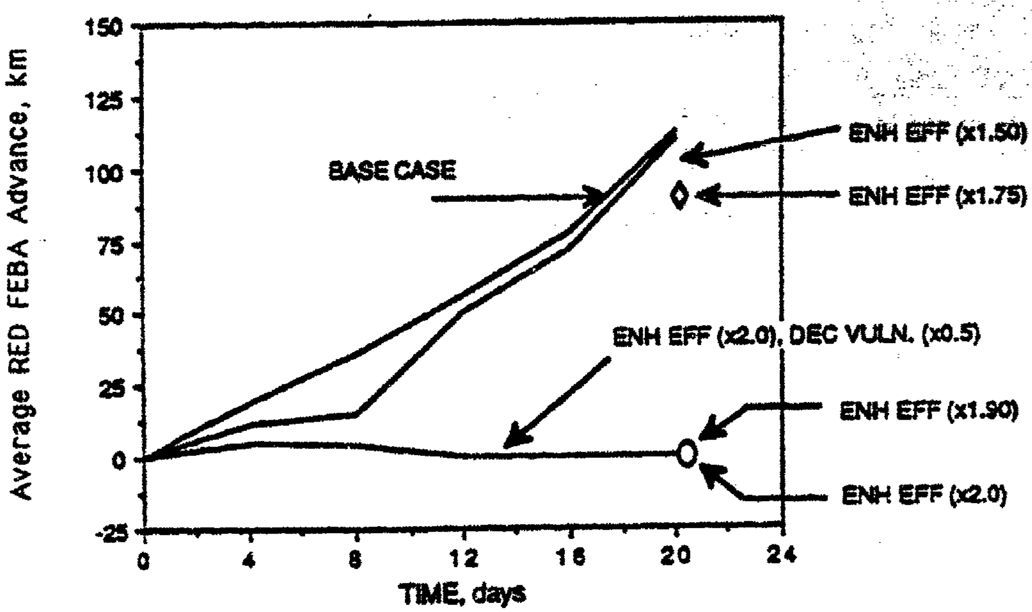


Figure 10. Effect of Separate Variations in Tank Effectiveness (Enhanced) and Tank Vulnerability (Reduced)

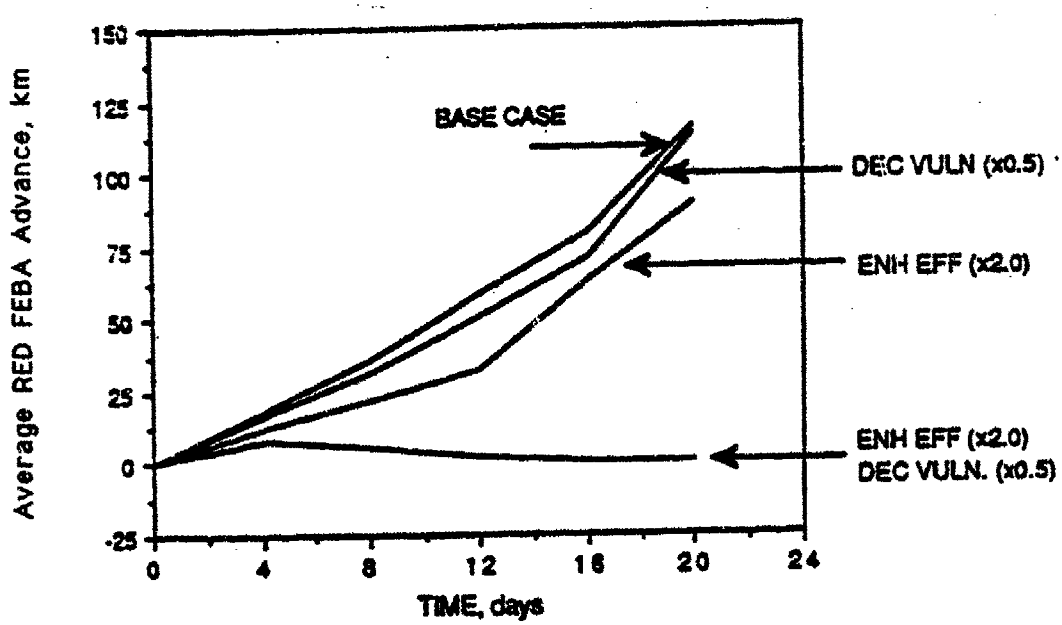


Figure 11. Effect of Separate Variations in Air Attack Effectiveness and Attack Aircraft Vulnerability

The factor of two experimental improvement in bombing accuracy for the A-10 or the F-16 was achieved in shallow dive bombing, a tactic that maximizes exposure to the defenses. It is not known whether the additional factor of 2 in effectiveness that would be achieved by shifting to the F-16 from the A-10 aircraft would have an effect on the war equivalent to achieving a factor of two in added survivability; it might not, because survivability is reflected in sorties flown, which impacts the outcome of the war differently than pure effectiveness change. Exposure to defenses could be reduced through level radar bombing or toss bombing, and under the appropriate circumstances the combination of effectiveness and survivability improvement in that kind of bombing could be achieved through some combination of training and avionics change or total aircraft system change. However, radar and toss bombing are significantly less accurate than day-visual dive bombing, so that gains in survivability might come at the operational cost of reduced overall effectiveness in bombing.

As in the tank case, training would be required with any of the equipment, and there is a suggestion in the TACAIR data examined, as for the tank case, that the training requirement with the new equipment could be reduced if the new equipment had more automatic features than the old.

2. Research Question 2: Hardware-Training Trade-offs

> 2. Is it feasible to trade off expenditures for training against those for hardware designed to achieve similar effects in combat, i.e., to improve force capability?

The previous discussion, for both the individual and the combined cases, has already indicated much of what can be learned from the present exploration about the trade-offs between hardware and training to improve unit performance. Some further insights can be drawn from looking at those results in the aggregate.

Overall, it appears from the combination of both the tank and the aircraft results examined thus far (and the caveats about the quality and completeness of the data must always be kept in view) that the magnitudes of improvement in unit performance that the TACWAR model indicated would be necessary to reverse the course of the war will be difficult to achieve through training alone. Hardware improvement and training together will be needed, and even then it will be difficult. The needed factor improvements appear to be there potentially, but it must be noted that the other side will also have such improvements available to it from similar sources, so that gains by one side will be offset by comparable gains, of unknown magnitude, on the other.

Also, none of this bears on force size as yet; force size may be the largest factor affecting the outcome of the war in a model like TACWAR with the improvements in unit and equipment effectiveness that the current exploration suggests may be achievable. But the cost structure of force size changes for similar effectiveness changes may be quite different from those examined here, depending on the kind of force under consideration. Force size was not considered in this analysis, but it must be evaluated in this context eventually.

A further observation engendered by these results, obvious when it is stated but not so obvious *a priori*, is that training is needed with new as well as with existing equipment, and this changes the way the equipment/training trade-off question must be formulated. The issue is not whether, as was stated in Deitchman (1988), funds at the margin should be put into improvement in training or equipment. Both contribute to force improvement, and both are needed.

The proper way to view the training/equipment trade-off at the margin is to break it into parts:

- First, to ascertain how much training will be needed to maximize performance with either current or new equipment, and
- Next, to decide at what point training has carried the force as far as it can go, so that equipment and force size change will be necessary to carry it further.

Funding must then be allocated among the different purposes. The trade-off to make at the margin is thus to allocate enough resources to training to make the best of the existing forces, and then to allocate funds to improve the forces' equipment and/or to change their size. More often, we allocate funds the other way: we improve equipment on some regular renewal cycle; we change force size when driven by external events; and we allocate funds to training from the residual if they are available.

Actually, although the above steps present the trade-offs in outline in terms of the stark parameters, the trade-offs must be followed through multiple system and force design and operating points until a satisfactory mix of training, hardware and force size expenditures is reached. Hopefully, the variation of capability with expenditure at the margins will be flat enough to permit flexibility in resource allocation, since not enough is likely to be known about the interaction among the training, the hardware and the force size effects to permit a firm "optimum" to be sought. The "optimum" would probably vary with specific equipment and type of warfare, in any case.

The current analysis leads to some insights that can inform the trade-off decision, particularly in helping to decide the level of resources required in each case; given the nature of the data, these must as yet be considered hypotheses to be tested:

- First, that the improvements achievable either with system upgrades or with enhanced training, in the armor and the TACAIR attack areas, may in many cases be of roughly comparable magnitude;
- Second, that the enhanced training and the equipment changes may in particular cases be of comparable cost;
- Third, that in some cases (e.g., a hypothetcal improvement of bombing avionics, in the present case) the equipment change may be of significantly lesser cost than the enhanced training to achieve the same result, and that the potential for such gains should be sought out on a case-by-case basis;
- Fourth, that more automatic modes in new systems may well reduce the requirement for individual proficiency training, freeing resources for more unit training;
- Fifth, that the use of networked, manned simulators in some combination with field training can significantly reduce the cost of unit training; and
- Sixth, that the capability of larger forces that are not as well trained should be compared with that of smaller forces that are better trained, in models like TACWAR, to complete the assessment of the most effective ways to spend resources to improve military capability overall.

An additional observation, related to the issue of equipment renewal, is that one cannot compare the improvements from equipment and training directly when the two achieve different things. Thus, for example, if a change in bombing avionics or tank subsystems allows fighting at night where that was not possible before, then no amount of training can provide that capability and there is little point in seeking the results from training that the new equipment can offer. On the other hand, fighting at night might be made possible by simple equipment augmentations, like flares over the battlefield. In that case, significant investment in training may be necessary to capitalize on the cheap equipment extension. This is simply to say that training/equipment trade-offs must be made in a context within which both fit, and that unique contributions of each must be accounted for separately.

A further point is that improvement in force capability derived from training and from equipment improvement are obtained on different time scales. For the short term (i.e., a year or two), training may be the only available source of improvement. Except under crash conditions in wartime (and often, even then), depending on the circumstances and magnitude of the change, equipment improvements can take five years or more for subsystem changes, and up to 12 to 20 years for major system changes. Force size increases will have still different time scales of magnitude between those of increased training for existing forces and reequipping those forces. The time to expand forces is composed of the time necessary to recruit personnel, procure their equipment, and train them in the use of the equipment and in operation as forces. All this indicates that time scales of change must also enter the algorithms encompassing the three sources of force improvement.

V. CONCLUSIONS AND DIRECTIONS FOR FUTURE EXPLORATION

A. OVERALL CONCLUSIONS

The following conclusions may be drawn about the problem of quantifying the military value of training for system acquisition and resource allocation decision purposes:

1. It appears possible to quantify the military value of training, but data bearing directly on the relationships of interest will have to be gathered through trials designed explicitly for the purpose. The duration of the trials should be long enough to indicate the asymptotic value of capability achievable through training. Resource availability may determine the thoroughness of the feasible experimentation, however, so that ad hoc opportunities to take advantage of existing data should not be foregone.

2. Experimental data about the impact of training on unit effectiveness gathered under controlled conditions in simulator networks like SIMNET will be useful in quantifying the military value of training, and they will be better controlled and less expensive than field exercises. Results from field exercises will offer insights into some aspects of unit operational training that may not be available from the simulator environment. Assessment of both kinds of results will involve speculative elements based on qualitative judgment as well as quantitative comparisons.

3. Some of the data suggest that more automatic modes in new systems may reduce the requirement for individual proficiency training, freeing resources for more unit training. The experimental data needed to quantify the military value of training must include attention to the interactions of equipment design with unit proficiency, including especially the impact, if any, of extensive system automation on *unit*, as distinguished from individual, training effectiveness.

4. It will not be easy to achieve the degree of improvement in unit capability that the initial exploration in this series indicated would be necessary to reverse the negative outcome of the NATO Central Region war modeled. Elements of training and equipment improvement and replacement will have to be combined to have any chance of achieving such results.

5. The cases explored suggest that either training or equipment improvement for specific military tasks produce effects on force effectiveness that are roughly

comparable in magnitude. Depending on the cost elements included, training or equipment improvement may be either comparable in cost or else training tends to cost less--sometimes considerably less. Other cases than those examined here may show a different balance. In addition, force size changes to achieve similar effects and the resources that would be required would have to figure in such assessments. Regardless of how the cost and performance comparisons may vary when explored in more detail, it is apparent that algorithms for allocation of resources among training, equipment improvement and force size must be devised to seek the most efficient resource allocation among the available approaches to force improvement. Such algorithms are not currently used in cost-effectiveness analyses of new systems.

B. NEXT STEPS

The next steps in exploration of the military value of training follow directly from the discussion and conclusions above. In all cases, the areas of armored combat and tactical air-to-ground attack should continue to be the ones explored. There are several reasons for this. The first is that the earlier analysis of military value showed that these two areas hold the key to major changes in the outcome of a battle or a war. The second is that the Services, recognizing that fact, devote a great deal of training and equipment resources to these areas of military activity. This includes experimental activity involving both field exercises, which are now coming to have large components of quantitative measurement, and simulation with more than single platform units. This, in turn, generates experimental data that at least implicitly show how unit effectiveness changes with practice and that might be exploited for analytical purposes. Finally, the resource expenditures in the two areas combined represent a significant enough fraction of the general purpose forces budget that any useful results can have a large impact on how that budget is spent.

Within these two areas, the outcome of the current explorations suggests the following next steps for further exploration into the problem of quantifying the military value of training:

- A much more thorough data exploration, including but not limited to past work at SIMNET and field exercises like those at the NTC and at Red Flag, (a) to define the nature of the experimental data available, and the problems of access to and manipulation of the data for analytical purposes, and (b) to see what more can be learned about the effects of exercises and training on unit performance;[11]

- Enlist the Services' interest and help in designing and carrying out relevant trials at SIMNET and available, analogous USAF and USN simulator complexes, to shed light on the ultimate capabilities training can develop with specific equipment levels and types, on the feasibility of extending results from "pure" to combined arms units and from small to larger units, and to explore training-equipment interactions with special attention to the impact of automated combat subsystems in individual platforms on unit performance.
- These trials should also explore whether and how exercise and simulation data gathered at low levels of military organization, such as platoon or flight levels, aggregate to describe performance of units at higher levels of organization, such as battalion or squadron levels.
- Use the data gathered above to devise resource allocation algorithms incorporating both training and equipment effectiveness parameters.
- Experiment with the algorithms using warfare simulation models such as are used by the DoD for budget planning and evaluation purposes, to explore the algorithms' ranges of utility and how they might affect resource allocation in the Department of Defense. These explorations would include subjecting the results to military judgment, to ascertain whether the aggregation from organization levels such as platoons and battalions to divisions and armies appears to give reasonable results.

---

11 In this connection, ongoing work that was not in sufficiently well-developed form to enter this analysis has been brought to the author's attention during the review of this report (S. Horowitz, private communication). The work, which is likely to be available to inform subsequent analyses along the same lines, includes ongoing evaluations of SAC crew proficiency improvement with flying hours; data and analyses beginning to emerge from the NTC; and analyses, being performed by Hammon and Horowitz based on Navy data, of the relationship between aircrew flying hours and fighter kills in aerial combat.

REFERENCES

Army Institute for the Behavioral and Social Sciences, (ARI), "Initial Validation of REALTRAIN with Army Combat Units in Europe," October 1976.

BBN Systems and Technologies Corporation: Unpublished data from tests that entered Headquarters, Department of The Army, U.S.A. Armor Center, Fort Knox, Kentucky, "Abrams Tank Block II Product Improvement Package SIMNET Evaluation," November 1988; Data furnished, Spring, 1989.

Cedel, LTC E.C., and Fuchs, LTC R.P., *An Analysis of Factors Affecting Pilot Proficiency*, USAF Center for Studies and Analyses: Fighter Division Study Report, December 1986.

Deitchman, S.J., "Preliminary Exploration of the Use of a Warfare Simulation Model to Examine the Military Value of Training." IDA Paper P-2094, March 1988.

Graves, J.W., et al., "Main Battle Tank Study (U)," IDA Report R-203, October 1974 (SECRET).[12]

Hammon, C.P., and Horowitz, S.A., "Flying Hours and Crew Performance." IDA Working Paper, June 1989.

Headquarters, Department of The Army, U.S.A. Armor Center, Fort Knox, Kentucky, "Abrams Tank Block II Product Improvement Package SIMNET Evaluation," November 1988.

Joint Munitions Effectiveness Manual, (Air-to-Surface/Accuracy) (U). Book 2. Rev. Ed. 1985. Report No. TH61A1-3-3. Joint Technical Coordinating Group for Munitions Effectiveness. Tucker AFB, OK (CONFIDENTIAL).[13]

Kerlin, E.P., "The IDA Tactical Warfare Model: A Theater-Level Model of Conventional Nuclear and Chemical Warfare," IDA Report R-211, November 1977.

Lane, N., (Essex Corporation), "Skill Acquisition Curves and Military Training," IDA Paper P-1945, January 1986.

Mairs, L., et al., "Quantifying the Training to Readiness Links: The Case for Attack Pilots," SAG Corporation, November 1986.

Orlansky, J., and Thorpe, Col. J., *SIMNET-Engagement Training for Tactical Warfare* (U): to be published, Journal of Defense Research, 1990.

United State Military Academy, Department of Social Science: "Are Smart Tankers Better Tankers," AFQT and Military Productivity, December 1984.

[12, 13] No portion of this reference is quoted in the present report.